Score, pseudo-score and residual diagnostics for goodness-of-fit of spatial point process models

Baddeley, Adrian ; Rubak, Ege Holger; Møller, Jesper

Publication date:
2010

Document Version
Early version, also known as pre-print

AALBORG UNIVERSITY

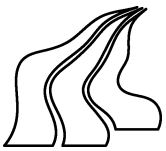
Score, pseudo-score and residual diagnostics for goodness-of-fit of spatial point process models

by

Adrian Baddeley, Ege Rubak and Jesper Møller

R-2010-06 August 2010

Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7 G ▪ DK - 9220 Aalborg Øst ▪ Denmark
Phone: +45 99 40 80 80 ▪ Telefax: +45 98 15 81 29
URL: http://www.math.aau.dk



ISSN 1399–2503 ▪ On-line version ISSN 1601–7811

Submitted to the Statistical Science

Score, pseudo-score and residual diagnostics for goodness-of-fit of spatial point process models

Adrian Baddeley, Ege Rubak and Jesper Møller

CSIRO and Aalborg University

Abstract. We develop new tools for formal inference and informal model validation in the analysis of spatial point pattern data. The score test is generalised to a 'pseudo-score' test derived from Besag's pseudo-likelihood, and to a class of diagnostics based on point process residuals. The results lend theoretical support to the established practice of using functional summary statistics such as Ripley's K-function, when testing for complete spatial randomness; and they provide new tools such as the *compensator* of the K-function for testing other fitted models. The results also support localisation methods such as the scan statistic and smoothed residual plots. Software for computing the diagnostics is provided.

AMS 2000 subject classifications: Primary 62M30; secondary 62J20.
Key words and phrases: compensator, functional summary statistics, model validation, point process residuals, pseudo-likelihood.

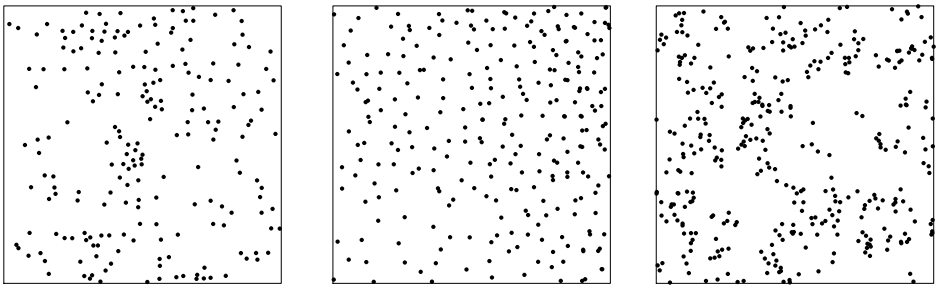
CSIRO Mathematics, Informatics and Statistics, Private Bag 5, Wembley WA 6913, Australia. (e-mail: Adrian.Baddeley@csiro.au). Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg Ø, Denmark (e-mail: rubak@math.aau.dk; jm@math.aau.dk).

1. INTRODUCTION

This paper develops new tools for formal inference and informal model validation in the analysis of spatial point pattern data. The score test statistic, based on the point process likelihood, is generalised to a 'pseudo-score' test statistic derived from Besag's pseudo-likelihood. The score and pseudo-score can be viewed as residuals, and further generalised to a class of residual diagnostics.

The likelihood score and the score test [73, 60], [21, pp 315 and 324] are used frequently in applied statistics to provide diagnostics for model selection and model validation [2, 18, 59, 15, 75]. In spatial statistics, the score test is used mainly to support formal inference about covariate effects [13, 46, 74] assuming the underlying point process is Poisson under both the null and alternative hypotheses. Our approach extends this to a much wider class of point processes, making it possible (for example) to check for covariate effects or localised hotspots in a clustered point pattern.

Figure 1 shows three example datasets studied in the paper. Our techniques make it possible to check separately for 'inhomogeneity' (spatial variation in abundance of points) and 'interaction' (localised dependence between points) in these data.



(a) (b) (c)

Fig 1: Point pattern datasets. (a) Japanese black pine seedlings and saplings in a 10×10 metre quadrat [52, 53]. Reprinted by kind permission of Professors M. Numata and Y. Ogata. (b) Simulated realisation of inhomogeneous Strauss process showing strong inhibition and spatial trend [6, Fig. 4b]. (c) Simulated realisation of homogeneous Geyer saturation process showing moderately strong clustering without spatial trend [6, Fig. 4c].

Our approach also provides theoretical support for the established practice of using functional summary statistics such as Ripley's K-function [62, 63] to study clustering and inhibition between points. In one class of models, the score test statistic is equivalent to the empirical K-function, and the score test procedure is closely related to the customary goodness-of-fit procedure based on comparing the empirical K-function with its null expected value. Similar statements apply to the nearest neighbour distance distribution function G and the empty space function F.

For computational efficiency, especially in large datasets, the point process likelihood is often replaced by Besag's [14] pseudo-likelihood. The resulting 'pseudo-score' is a possible surrogate for the likelihood score. In one model, the pseudo-score test statistic is equivalent to a *residual* version of the empirical

K-function, yielding a new, efficient diagnostic for goodness-of-fit. However, in general, the interpretation of the pseudo-score test statistic is conceptually more complicated than that of the likelihood score test statistic, and hence difficult to employ as a diagnostic.

In classical settings the score test statistic is a weighted sum of residuals. Here the pseudo-score test statistic is a weighted point process residual in the sense of [6, 3]. This suggests a simplification, in which the pseudo-score test statistic is replaced by another residual diagnostic that is easier to interpret and to compute.

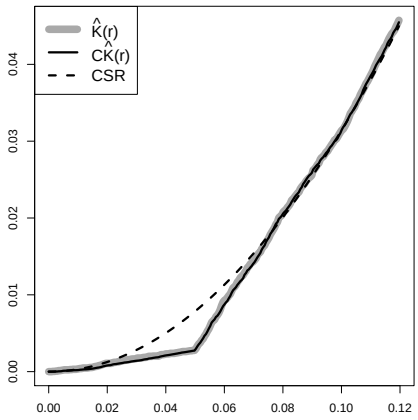


Fig 2: Empirical K-function (thick grey line) for the point pattern data in Figure 1b, compensator of the K-function (solid black line) for a model of the correct form, and expected K-function for a homogeneous Poisson process (dashed line).

In special cases this diagnostic is a residual version of one of the classical functional summary statistics K, G or F obtained by subtracting a *'compensator'* from the functional summary statistic. The compensator depends on the observed data and on the fitted model. For example, if the fitted model is the homogeneous Poisson process, then the compensator of $G(r)$ is $F(r)$, and the compensator of $K(r)$ is πr^2. This approach provides a new class of residual summary statistics that can be used as informal diagnostics for goodness-of-fit, for a wide range of point process models, in close analogy with current practice. The diagnostics apply under very general conditions, including the case of inhomogeneous point process models, where exploratory methods are underdeveloped or inapplicable. For instance, the compensator of $K(r)$ for an inhomogeneous non-Poisson model is illustrated in Figure 2.

Section 2 introduces basic definitions and assumptions. Section 3 describes the score test for a general point process model, and Section 4 develops the important case of Poisson point process models. Section 5 gives examples and technical tools for non-Poisson point process models. Section 6 develops the general theory for our diagnostic tools. Section 7 applies these tools to tests for first order trend and hotspots. Sections 8–11 develop diagnostics for interaction between points, based on pairwise distances, nearest neighbour distances and empty space distances respectively. The tools are demonstrated on data in Sections 12–

15. Further examples of diagnostics are given in Appendix A. Appendices B–E provide technical details.

2. ASSUMPTIONS

2.1 Fundamentals

A spatial point pattern dataset is a finite set $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ of points $x_i \in W$, where the number of points $n(\boldsymbol{x}) = n \geq 0$ is not fixed in advance, and the domain of observation $W \subset \mathbb{R}^d$ is a fixed, known region of d-dimensional space with finite positive volume $|W|$. We take $d = 2$ but the results generalise easily to all dimensions.

A point process model assumes that $\boldsymbol{x}$ is a realisation of a finite point process $\boldsymbol{X}$ in W without multiple points. We can equivalently view $\boldsymbol{X}$ as a random finite subset of W. Much of the literature on spatial statistics assumes that $\boldsymbol{X}$ is the restriction $\boldsymbol{X} = \boldsymbol{Y} \cap W$ of a stationary point process $\boldsymbol{Y}$ on the entire space $\mathbb{R}^2$. We do not assume this; there is no assumption of stationarity, and some of the models considered here are intrinsically confined to the domain W. For further background material including measure theoretical details, see e.g. [50, Appendix B].

Write $\boldsymbol{X} \sim \mathrm{Poisson}(W, \rho)$ if $\boldsymbol{X}$ follows the Poisson process on W with intensity function ρ, where we assume $\nu = \int_W \rho(u)\,\mathrm{d}u$ is finite. Then $n(\boldsymbol{X})$ is Poisson distributed with mean ν, and conditional on $n(\boldsymbol{X})$, the points in $\boldsymbol{X}$ are i.i.d. with density $\rho(u)/\nu$.

Every point process model considered here is assumed to have a probability density with respect to $\mathrm{Poisson}(W, 1)$, the unit rate Poisson process, under one of the following scenarios.

2.2 Unconditional case

In the *unconditional case* we assume $\boldsymbol{X}$ has a density f with respect to $\mathrm{Poisson}(W, 1)$. Then the density is characterised by the property

$$\mathbb{E}[h(\boldsymbol{X})] = \mathbb{E}[h(\boldsymbol{Y}) f(\boldsymbol{Y})] \tag{1}$$

for all non-negative measurable functionals h, where $\boldsymbol{Y} \sim \mathrm{Poisson}(W, 1)$. In particular the density of $\mathrm{Poisson}(W, \rho)$ is

$$f(\boldsymbol{x}) = \exp\left(\int_W (1 - \rho(u))\,\mathrm{d}u\right) \prod_i \rho(x_i). \tag{2}$$

We assume that f is hereditary, i.e. $f(\boldsymbol{x}) > 0$ implies $f(\boldsymbol{y}) > 0$ for all finite $\boldsymbol{y} \subset \boldsymbol{x} \subset W$.

2.3 Conditional case

In the *conditional case*, we assume $\boldsymbol{X} = \boldsymbol{Y} \cap W$ where $\boldsymbol{Y}$ is a point process. Thus $\boldsymbol{X}$ may depend on unobserved points of $\boldsymbol{Y}$ lying outside W. The density of $\boldsymbol{X}$ may be unknown or intractable. Under suitable conditions (explained in Section 5.4) modelling and inference can be based on the conditional distribution of $\boldsymbol{X}^\circ = \boldsymbol{X} \cap W^\circ$ given $\boldsymbol{X}^+ = \boldsymbol{X} \cap W^+ = \boldsymbol{x}^+$, where $W^+ \subset W$ is a subregion, typically a region near the boundary of W, and only the points in

$W^\circ = W \setminus W^+$ are treated as random. We assume that the conditional distribution of $\boldsymbol{X}^\circ = \boldsymbol{X} \cap W^\circ$ given $\boldsymbol{X}^+ = \boldsymbol{X} \cap W^+ = \boldsymbol{x}^+$ has an hereditary density $f(\boldsymbol{x}^\circ \mid \boldsymbol{x}^+)$ with respect to $\mathrm{Poisson}(W^\circ, 1)$.

For ease of exposition, we focus mainly on the unconditional case, with occasional comments on the conditional case. For Poisson point process models, we always take $W = W^\circ$ so that the two cases agree.

3. SCORE TEST FOR POINT PROCESSES

In principle, any technique for likelihood-based inference is applicable to point process likelihoods. In practice, many likelihood-based computations require extensive Monte Carlo simulation [30, 50, 49]. To minimise such difficulties, when assessing the goodness-of-fit of a fitted point process model, it is natural to choose the score test which only requires computations for the null hypothesis [73, 60].

Consider any parametric family of point process models for $\boldsymbol{X}$ with density f_θ indexed by a k-dimensional vector parameter $\theta \in \Theta \subseteq \mathbb{R}^k$. For a *simple* null hypothesis $H_0 : \theta = \theta_0$ where $\theta_0 \in \Theta$ is fixed, the score test against any alternative $H_1 : \theta \in \Theta_1$, where $\Theta_1 \subseteq \Theta \setminus \{\theta_0\}$, is based on the score test statistic [21, p. 315]

$$T^2 = U(\theta_0)^{\mathrm{T}} I(\theta_0)^{-1} U(\theta_0). \tag{3}$$

Here $U(\theta) = \frac{\partial}{\partial \theta} \log f_\theta(\boldsymbol{x})$ and $I(\theta) = \mathbb{E}_\theta \left[U(\theta) U(\theta)^{\mathrm{T}} \right]$ are the score function and Fisher information respectively, and the expectation is with respect to f_θ. Here and throughout, we assume that the order of integration and differentation *with respect to* θ can be interchanged. Under suitable conditions, the null distribution of T^2 is χ^2 with k degrees of freedom. In the case $k = 1$ it may be informative to evaluate the signed square root

$$T = U(\theta_0) / \sqrt{I(\theta_0)} \tag{4}$$

which is asymptotically standard normally distributed under the same conditions.

For a *composite* null hypothesis $H_0 : \theta \in \Theta_0$ where $\Theta_0 \subset \Theta$ is an m-dimensional submanifold with $0 < m < k$, the score test statistic is defined in [21, p. 324]. However, we shall not use this version of the score test, as it assumes differentiability of the likelihood with respect to nuisance parameters, which is not necessarily applicable here (as exemplified in Section 4.2).

In the sequel we often consider models of the form

$$f_{(\alpha,\beta)}(\boldsymbol{x}) = c(\alpha, \beta) h_\alpha(\boldsymbol{x}) \exp(\beta S(\boldsymbol{x})) \tag{5}$$

where the parameter β and the statistic $S(\boldsymbol{x})$ are one dimensional, and the null hypothesis is $H_0 : \beta = 0$. For fixed α, this is a linear exponential family and (4) becomes

$$T(\alpha) = \big(S(\boldsymbol{x}) - \mathbb{E}_{(\alpha,0)}[S(\boldsymbol{x})] \big) \, / \sqrt{\mathbb{V}\mathrm{ar}_{(\alpha,0)}[S(\boldsymbol{x})]}.$$

In practice, when α is unknown, we replace α by its MLE under H_0 so that, with a slight abuse of notation, the signed square root of the score test statistic is approximated by

$$T = T(\hat{\alpha}) = \big(S(\boldsymbol{x}) - \mathbb{E}_{(\hat{\alpha},0)}[S(\boldsymbol{x})] \big) \, / \sqrt{\mathbb{V}\mathrm{ar}_{(\hat{\alpha},0)}[S(\boldsymbol{x})]}. \tag{6}$$

Under suitable conditions, T in (6) is asymptotically equivalent to T in (4), and so a standard Normal approximation may still apply.

4. SCORE TEST FOR POISSON PROCESSES

Application of the score test to Poisson point process models appears to originate with Cox [20]. Consider a parametric family of Poisson processes, $\mathrm{Poisson}(W, \rho_\theta)$, where the intensity function is indexed by $\theta \in \Theta$. The score test statistic is (3) where

$$\begin{aligned} U(\theta) &= \sum_i \kappa_\theta(x_i) - \int_W \kappa_\theta(u)\rho_\theta(u)\,\mathrm{d}u \\ I(\theta) &= \int_W \kappa_\theta(u)\,\kappa_\theta(u)^{\mathrm{T}}\rho_\theta(u)\,\mathrm{d}u \end{aligned}$$

with $\kappa_\theta(u) = \frac{\partial}{\partial\theta}\log\rho_\theta(u)$. Asymptotic results are given in [44, 61].

4.1 Log-linear alternative

The score test is commonly used in spatial epidemiology to assess whether disease incidence depends on environmental exposure. As a particular case of (5), suppose the Poisson model has a log-linear intensity function

$$\rho_{(\alpha,\beta)}(u) = \exp(\alpha + \beta Z(u)) \tag{7}$$

where $Z(u), u \in W$ is a known, real-valued and non-constant covariate function, and α and β are real parameters. Cox [20] noted that the uniformly most powerful test of $H_0 : \beta = 0$ (the homogeneous Poisson process) against $H_1 : \beta > 0$ is based on the statistic

$$S(\boldsymbol{x}) = \sum_i Z(x_i). \tag{8}$$

Recall that, for a point process $\boldsymbol{X}$ on W with intensity function ρ,

$$\mathbb{E}\left(\sum_{x_i\in\boldsymbol{X}} h(x_i)\right) = \int_W h(u)\rho(u)\,\mathrm{d}u \tag{9}$$

for any Borel function h such that the integral on the right hand side exists, and for $\mathrm{Poisson}(W, \rho)$,

$$\mathbb{V}\mathrm{ar}\left(\sum_{x_i\in\boldsymbol{X}} h(x_i)\right) = \int_W h(u)^2\rho(u)\,\mathrm{d}u \tag{10}$$

for any Borel function h such that the integral on the right hand side exists [23, p. 188]. Hence the standardised version of (8) is

$$T = \left(S(\boldsymbol{x}) - \hat\kappa\int_W Z(u)\,\mathrm{d}u\right) \Big/ \sqrt{\hat\kappa\int_W Z(u)^2\,\mathrm{d}u} \tag{11}$$

where $\hat\kappa = n/|W|$ is the MLE of the intensity $\kappa = \exp(\alpha)$ under the null hypothesis. This is a direct application of the approximation (6) of the signed square root of the score test statistic.

Berman [13] proposed several tests and diagnostics for spatial association between a point process $\boldsymbol{X}$ and a covariate function $Z(u)$. Berman's Z_1 test is equivalent to the Cox score test described above. Waller *et al*. [74] and Lawson [46] proposed tests for the dependence of disease incidence on environmental exposure, based on data giving point locations of disease cases. These are also applications of the score test. Berman conditioned on the number of points when making inference. This is in accordance with the observation that the statistic $n(\boldsymbol{x})$ is S-ancillary for β, while $S(\boldsymbol{x})$ is S-sufficient for β.

4.2 Threshold alternative and nuisance parameters

Consider the Poisson process with an intensity function of 'threshold' form,

$$\rho_{z,\kappa,\phi}(u) = \begin{cases} \kappa \exp(\phi) & \text{if } Z(u) \le z \\ \kappa & \text{if } Z(u) > z \end{cases}$$

where z is the threshold level. If z is fixed, this model is a special case of (7) with $Z(u)$ replaced by $\mathbb{I}\{Z(u) \le z\}$, and so (8) is replaced by

$$S(\boldsymbol{x}) = S(\boldsymbol{x}, z) = \sum_i \mathbb{I}\{Z(x_i) \le z\}$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function. By (11) the (approximate) score test of $H_0 : \phi = 0$ against $H_1 : \phi \ne 0$ is based on

$$T = T(z) = \left(S(\boldsymbol{x}, z) - \hat{\kappa} A(z)\right) / \sqrt{\hat{\kappa} A(z)}$$

where $A(z) = |\{u \in W : \, Z(u) \le z\}|$ is the area of the corresponding level set of Z.

If z is not fixed, then it plays the role of a nuisance parameter in the score test: the value of z affects inference about the canonical parameter ϕ, which is the parameter of primary interest in the score test. Note that the likelihood is not differentiable with respect to z.

In most applications of the score test, a nuisance parameter would be replaced by its MLE under the null hypothesis. However in this context, z is not identifiable under the null hypothesis. Several approaches to this problem have been proposed [17, 24, 25, 32, 67]. They include replacing z by its MLE under the alternative [17], maximising $T(z)$ or $|T(z)|$ over z [24, 25], and finding the maximum p-value of $T(z)$ or $|T(z)|$ over a confidence region for z under the alternative [67].

These approaches appear to be inapplicable to the current context. While the null distribution of $T(z)$ is asymptotically $N(0, 1)$ for each fixed z as $\kappa \to \infty$, this convergence is not uniform in z. The null distribution of $S(\boldsymbol{x}, z)$ is Poisson with parameter $\kappa A(z)$; sample paths of $T(z)$ will be governed by Poisson behaviour where $A(z)$ is small.

In this paper, our approach is simply to plot the score test statistic as a function of the nuisance parameter. This turns the score test into a graphical exploratory tool, following the approach adopted in many other areas [2, 18, 59, 15, 75]. A second style of plot based on $S(x, z) - \hat{\kappa} A(z)$ against z may be more appropriate visually. Such a plot is the lurking variable plot of [6]. Berman [13] also proposed a plot of $S(\boldsymbol{x}, z)$ against z, together with a plot of $\hat{\kappa} A(z)$ against z, as

a diagnostic for dependence on Z. This is related to the Kolmogorov-Smirnov test since, under H_0, the values $Y_i = Z(x_i)$ are i.i.d. with distribution function $\mathbb{P}(Y \le y) = A(z)/|W|$.

4.3 Hot spot alternative

Consider the Poisson process with intensity

$$\rho_{\kappa,\phi,v}(u) = \kappa \exp(\phi k(u-v)) \tag{12}$$

where k is a kernel (a probability density on $\mathbb{R}^2$), $\kappa > 0$ and ϕ are real parameters, and $v \in \mathbb{R}^2$ is a nuisance parameter. This process has a 'hot spot' of elevated intensity in the vicinity of the location v. By (11) and (9)–(10) the score test of $H_0 : \phi = 0$ against $H_1 : \phi \neq 0$ is based on

$$T = T(v) = (S(\boldsymbol{x}, v) - \hat{\kappa} M_1(v))/\sqrt{\hat{\kappa} M_2(v)}$$

where

$$S(\boldsymbol{x}, v) = \sum_i k(x_i - v)$$

is the usual nonparametric kernel estimate of point process intensity [27] evaluated at v without edge correction, and

$$M_i(v) = \int_W k(u-v)^i \, \mathrm{d}u, \quad i = 1, 2.$$

The numerator $S(\boldsymbol{x}, v) - \hat{\kappa} M_1(v)$ is the *smoothed residual field* [6] of the null model. In the special case where $k(u) \propto \mathbb{I}\{\|u\| \le h\}$ is the uniform density on a disc of radius h, the maximum $\max_v T(v)$ is closely related to the *scan statistic* [1, 43].

5. NON-POISSON MODELS

The remainder of the paper deals with the case where the alternative (and perhaps also the null) is not a Poisson process. Key examples are stated in Section 5.1. Non-Poisson models require additional tools including the conditional intensity (Section 5.2) and pseudo-likelihood (Section 5.3).

5.1 Point process models with interaction

We shall frequently consider densities of the form

$$f(\boldsymbol{x}) = c \left[\prod_i \lambda(x_i) \right] \exp\left(\phi V(\boldsymbol{x})\right) \tag{13}$$

where c is a normalising constant, the first order term λ is a non-negative function, ϕ is a real interaction parameter, and $V(\boldsymbol{x})$ is a real non-additive function which specifies the interaction between the points. We refer to V as the interaction potential. In general, apart from the Poisson density (2) corresponding to the case $\phi = 0$, the normalising constant is not expressible in closed form.

Often the definition of V can be extended to all finite point patterns in $\mathbb{R}^2$ so as to be invariant under rigid motions (translations and rotations). Then the model for $\boldsymbol{X}$ is said to be homogeneous if λ is constant on W, and inhomogeneous otherwise.

Let

$$d(u, \boldsymbol{x}) = \min_j \|u - x_j\|$$

denote the distance from a location u to its nearest neighbour in the point configuration $\boldsymbol{x}$. For $n(\boldsymbol{x}) = n \geq 1$ and $i = 1, \ldots, n$, define

$$\boldsymbol{x}_{-i} = \boldsymbol{x} \setminus \{x_i\}.$$

In many places in this paper we consider the following three motion-invariant interaction potentials $V(\boldsymbol{x}) = V(\boldsymbol{x}, r)$ depending on a parameter $r > 0$ which specifies the range of interaction. The *Strauss process* [71] has interaction potential

$$V_S(\boldsymbol{x}, r) = \sum_{i<j} \mathbb{I}\{\|x_i - x_j\| \leq r\} \tag{14}$$

the number of r-close pairs of points in $\boldsymbol{x}$; the *Geyer saturation model* [30] with saturation threshold 1 has interaction potential

$$V_G(\boldsymbol{x}, r) = \sum_i \mathbb{I}\{d(x_i, \boldsymbol{x}_{-i}) \leq r\} \tag{15}$$

the number of points in $\boldsymbol{x}$ whose nearest neighbour is closer than r units; and the Widom-Rowlinson penetrable sphere model [76] or *area-interaction process* [11] has interaction potential

$$V_A(\boldsymbol{x}, r) = -|W \cap \bigcup_i B(x_i, r)| \tag{16}$$

the negative area of W intersected with the union of balls $B(x_i, r)$ of radius r centred at the points of $\boldsymbol{x}$. Each of these densities favours spatial clustering (positive association) when $\phi > 0$ and spatial inhibition (negative association) when $\phi < 0$. The Geyer and area-interaction models are well-defined point processes for any value of ϕ [11, 30], but the Strauss density is integrable only when $\phi \leq 0$ [42].

5.2 Conditional intensity

Consider a parametric model for a point process $\boldsymbol{X}$ in $\mathbb{R}^2$, with parameter $\theta \in \Theta$. Papangelou [58] defined the *conditional intensity* of $\boldsymbol{X}$ as a non-negative stochastic process $\lambda_\theta(u, \boldsymbol{X})$ indexed by locations $u \in \mathbb{R}^2$ and characterised by the property that

$$\mathbb{E}_\theta \left[\sum_{x_i \in \boldsymbol{X}} h(x_i, \boldsymbol{X} \setminus \{x_i\}) \right] = \mathbb{E}_\theta \left[\int_{\mathbb{R}^2} h(u, \boldsymbol{X}) \lambda_\theta(u, \boldsymbol{X}) \, \mathrm{d}u \right] \tag{17}$$

for all measurable functions h such that the left or right hand side exists. Equation (17) is known as the *Georgii-Nguyen-Zessin (GNZ) formula* [29, 40, 41, 51]; see also Section 6.4.1 in [50]. Adapting a term from stochastic process theory, we will call the random integral on the right side of (17) the *(Papangelou) compensator* of the random sum on the left side.

Consider a finite point process $\boldsymbol{X}$ in W. In the unconditional case (Section 2.2) we assume $\boldsymbol{X}$ has density $f_\theta(\boldsymbol{x})$ which is hereditary for all $\theta \in \Theta$. We may simply define

$$\lambda_\theta(u, \boldsymbol{x}) = f_\theta(\boldsymbol{x} \cup \{u\})/f_\theta(\boldsymbol{x}) \tag{18}$$

for all locations $u \in W$ and point configurations $\boldsymbol{x} \subset W$ such that $u \notin \boldsymbol{x}$. Here we take $0/0 = 0$. For $x_i \in \boldsymbol{x}$ we set $\lambda_\theta(x_i, \boldsymbol{x}) = \lambda_\theta(x_i, \boldsymbol{x}_{-i})$, and for $u \notin W$ we set $\lambda_\theta(u, \boldsymbol{x}) = 0$. Then it may be verified directly from (1) that (17) holds, so that (18) is the Papangelou conditional intensity of $\boldsymbol{X}$. Note that the normalising constant of f_θ cancels in (18). For a Poisson process, it follows from (2) and (18) that the conditional intensity is equivalent to the intensity function of the process.

In the conditional case (Section 2.3) we assume that the conditional distribution of $\boldsymbol{X}^\circ = \boldsymbol{X} \cap W^\circ$ given $\boldsymbol{X}^+ = \boldsymbol{X} \cap W^+ = \boldsymbol{x}^+$ has an hereditary density $f_\theta(\boldsymbol{x}^\circ \mid \boldsymbol{x}^+)$ with respect to $\mathrm{Poisson}(W^\circ, 1)$, for all $\theta \in \Theta$. Then define

$$\lambda_\theta(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+) = f_\theta(\boldsymbol{x}^\circ \cup \{u\} \mid \boldsymbol{x}^+)/f_\theta(\boldsymbol{x}^\circ \setminus \{u\} \mid \boldsymbol{x}^+) \tag{19}$$

if $u \in W^\circ$, and zero otherwise. It can similarly be verified that this is the Papangelou conditional intensity of the conditional distribution of $\boldsymbol{X}^\circ$ given $\boldsymbol{X}^+ = \boldsymbol{x}^+$.

It is convenient to rewrite (18) in the form

$$\lambda_\theta(u, \boldsymbol{x}) = \exp(\Delta_u \log f(\boldsymbol{x}))$$

where Δ is the one-point difference operator

$$\Delta_u h(\boldsymbol{x}) = h(\boldsymbol{x} \cup \{u\}) - h(\boldsymbol{x} \setminus \{u\}). \tag{20}$$

Note the Poincaré inequality for the Poisson process $\boldsymbol{X}$

$$\mathbb{V}\mathrm{ar}[h(\boldsymbol{X})] \le \mathbb{E} \int_W [\Delta_u h(\boldsymbol{X})]^2 \rho(u) \, \mathrm{d}u \tag{21}$$

holding for all measurable functionals h such that the right hand side is finite; see [45, 77].

5.3 Pseudo-likelihood and pseudo-score

To avoid computational problems with point process likelihoods, Besag [14] introduced the *pseudo-likelihood* function

$$\mathsf{PL}(\theta) = \left[\prod_i \lambda_\theta(x_i, \boldsymbol{x}) \right] \exp\left(- \int_W \lambda_\theta(u, \boldsymbol{x}) \, \mathrm{d}u \right). \tag{22}$$

This is of the same functional form as the likelihood function of a Poisson process (2), but has the conditional intensity in place of the Poisson intensity. The corresponding *pseudo-score*

$$\mathsf{PU}(\theta) = \frac{\partial}{\partial \theta} \log \mathsf{PL}(\theta) = \sum_i \frac{\partial}{\partial \theta} \log \lambda_\theta(x_i, \boldsymbol{x}) - \int_W \frac{\partial}{\partial \theta} \lambda_\theta(u, \boldsymbol{x}) \, \mathrm{d}u \tag{23}$$

is an unbiased estimating function (i.e. $PU(\theta)$ has zero-mean) by virtue of (17).

The pseudo-likelihood function can also be defined in the conditional case [38]. In (22) the product is instead over points $x_i \in \boldsymbol{x}^\circ$ and the integral is instead over W°; in (23) the sum is instead over points $x_i \in \boldsymbol{x}^\circ$ and the integral is instead over W°; and in both places $\boldsymbol{x} = \boldsymbol{x}^\circ \cup \boldsymbol{x}^+$. The conditional intensity $\lambda_\theta(u, \boldsymbol{x})$ must also be replaced by $\lambda_\theta(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)$.

5.4 Markov point processes

For a point process $\boldsymbol{X}$ constructed as $\boldsymbol{X} = \boldsymbol{Y} \cap W$ where $\boldsymbol{Y}$ is a point process in $\mathbb{R}^2$, the density and conditional intensity of $\boldsymbol{X}$ may not be available in simple form. Progress can be made if $\boldsymbol{Y}$ is a *Markov point process* of interaction range $R < \infty$, see [29, 51, 65, 72] and [50, Sect. 6.4.1]. Briefly, this means that the conditional intensity $\lambda_\theta(u, \boldsymbol{Y})$ of $\boldsymbol{Y}$ satisfies $\lambda_\theta(u, \boldsymbol{Y}) = \lambda_\theta(u, \boldsymbol{Y} \cap B(u, R))$ where $B(u, R)$ is the ball of radius R centred at u. Define the erosion of W by distance R

$$W_{\ominus R} = \{u \in W : \, B(u, R) \subset W\}$$

and assume this has non-zero area. Let $B = W \setminus W_{\ominus R}$ be the border region. The process satisfies a spatial Markov property: the processes $\boldsymbol{Y} \cap W_{\ominus R}$ and $\boldsymbol{Y} \cap W^c$ are conditionally independent given $\boldsymbol{Y} \cap B$.

In this situation we shall invoke the conditional case with $W^\circ = W_{\ominus R}$ and $W^+ = W \setminus W^\circ$. The conditional distribution of $\boldsymbol{X} \cap W^\circ$ given $\boldsymbol{X} \cap W^+ = \boldsymbol{x}^+$ has Papangelou conditional intensity

$$\lambda_\theta(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+) = \begin{cases} \lambda_\theta(u, \boldsymbol{x}^\circ \cup \boldsymbol{x}^+) & \text{if } u \in W^\circ \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

Thus *the unconditional and conditional versions of a Markov point process have the same Papangelou conditional intensity* at locations in W°.

For $\boldsymbol{x}^\circ = \{x_1, \ldots, x_{n^\circ}\}$, the conditional probability density becomes

$$f_\theta(\boldsymbol{x}^\circ \mid \boldsymbol{x}^+) = c_\theta(\boldsymbol{x}^+) \lambda_\theta(x_1, \boldsymbol{x}^\circ) \prod_{i=2}^{n^\circ} \lambda_\theta(x_i, \{x_1, \ldots, x_{i-1}\} \cup \boldsymbol{x}^+)$$

if $n^\circ > 0$, and $f_\theta(\emptyset \mid \boldsymbol{x}^+) = c_\theta(\boldsymbol{x}^+)$, where $\emptyset$ denotes the empty configuration, and the inverse normalising constant $c_\theta(\boldsymbol{x}^+)$ depends only on $\boldsymbol{x}^+$.

For example, instead of (13) we now consider

$$f(\boldsymbol{x}^\circ \mid \boldsymbol{x}^+) = c(\boldsymbol{x}^+) \left[\prod_{i=1}^{n^\circ} \lambda(x_i) \right] \exp\left(\phi V(\boldsymbol{x}^\circ \cup \boldsymbol{x}^+)\right)$$

assuming $V(\boldsymbol{y})$ is defined for all finite $\boldsymbol{y} \subset \mathbb{R}^2$ such that for any $u \in \mathbb{R}^2 \setminus \boldsymbol{y}$, $\Delta_u V(\boldsymbol{y})$ depends only on u and $\boldsymbol{y} \cap B(u, R)$. This condition is satisfied by the interaction potentials (14)-(16); note that the range of interaction is $R = r$ for the Strauss process, and $R = 2r$ for both the Geyer and the area-interaction models.

6. SCORE, PSEUDOSCORE AND RESIDUAL DIAGNOSTICS

This section develops the general theory for our diagnostic tools.

By (6) in Section 3 it is clear that comparison of a summary statistic $S(\boldsymbol{x})$ to its predicted value $\mathbb{E}S(\boldsymbol{X})$ under a null model, is effectively equivalent to the score test under an exponential family model where $S(\boldsymbol{x})$ is the canonical sufficient statistic. Similarly, the use of a *functional* summary statistic $S(\boldsymbol{x}, z)$, depending on a function argument z, is related to the score test under an exponential family model *where z is a nuisance parameter* and $S(\boldsymbol{x}, z)$ is the canonical sufficient statistic for fixed z. In this section we construct the corresponding exponential family models, apply the score test, and propose surrogates for the score test statistic.

6.1 Models

Let $f_\theta(\boldsymbol{x})$ be the density of any point process $\boldsymbol{X}$ on W governed by a parameter θ. Let $S(\boldsymbol{x}, z)$ be a functional summary statistic of the point pattern dataset $\boldsymbol{x}$, with function argument z belonging to any space.

Consider the *extended model* with density

$$f_{\theta,\phi,z}(\boldsymbol{x}) = c_{\theta,\phi,z} f_\theta(\boldsymbol{x}) \exp(\phi S(\boldsymbol{x}, z)) \tag{25}$$

where ϕ is a real parameter, and $c_{\theta,\phi,z}$ is the normalising constant. The density is well-defined provided

$$M(\theta, \phi, z) = \mathbb{E}\left[f_\theta(\boldsymbol{Y}) \exp(\phi S(\boldsymbol{Y}, z))\right] < \infty$$

where $\boldsymbol{Y} \sim \mathrm{Poisson}(W, 1)$. The extended model is constructed by 'exponential tilting' of the original model by the statistic S. By (6), for fixed θ and z, assuming differentiability of M with respect to ϕ in a neighbourhood of $\phi = 0$, the signed root of the score test statistic is approximated by

$$T = \left(S(\boldsymbol{x}, z) - \mathbb{E}_{\hat\theta}[S(\boldsymbol{X}, z)]\right) / \sqrt{\mathbb{V}\mathrm{ar}_{\hat\theta}[S(\boldsymbol{X}, z)]} \tag{26}$$

where $\hat\theta$ is the MLE under the null model, and the expectation and variance are with respect to the null model with density $f_{\hat\theta}$.

Insight into the qualitative behaviour of the extended model (25) can be obtained by studying the *perturbing model*

$$g_{\phi,z}(\boldsymbol{x}) = k_{\phi,z} \exp(\phi S(\boldsymbol{x}, z)), \tag{27}$$

provided this is a well-defined density with respect to $\mathrm{Poisson}(W, 1)$, where $k_{\phi,z}$ is the normalising constant. When the null hypothesis is a homogeneous Poisson process, the extended model is identical to the perturbing model, up to a change in the first order term. In general, the extended model is a qualitative hybrid between the null and perturbing models.

In this context the score test is equivalent to naive comparison of the observed and null-expected values of the functional summary statistic S. The test statistic T in (26) may be difficult to evaluate; typically, apart from Poisson models, the moments (particularly the variance) of S would not be available in closed form. The null distribution of T would also typically be unknown. Hence, implementation of the score test would typically require moment approximation and simulation from the null model, which in both cases may be computationally expensive. Various approximations for the score or the score test statistic can be constructed, as discussed in the sequel.

6.2 Pseudo-score of extended model

The extended model (25) is an exponential family with respect to ϕ, having conditional intensity

$$\kappa_{\theta,\phi,z}(u, \boldsymbol{x}) = \lambda_\theta(u, \boldsymbol{x}) \exp\left(\phi \Delta_u S(\boldsymbol{x}, z)\right)$$

where $\lambda_\theta(u, \boldsymbol{x})$ is the conditional intensity of the null model. The pseudo-score function with respect to ϕ, evaluated at $\phi = 0$, is

$$\mathsf{PU}(\theta, z) = \sum_i \Delta_{x_i} S(\boldsymbol{x}, z) - \int_W \Delta_u S(\boldsymbol{x}, z) \lambda_\theta(u, \boldsymbol{x}) \, \mathrm{d}u$$

where the first term

$$\Sigma\Delta\, S(\boldsymbol{x}, z) = \sum_i \Delta_{x_i} S(\boldsymbol{x}, z) \tag{28}$$

will be called the *pseudo-sum* of S. If $\hat{\theta}$ is the maximum pseudo-likelihood estimate (MPLE) under H_0, the second term with θ replaced by $\hat{\theta}$ becomes

$$\mathrm{C}\Delta\, S(\boldsymbol{x}, z) = \int_W \Delta_u S(\boldsymbol{x}, z) \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u \tag{29}$$

and will be called the *(estimated) pseudo-compensator* of S. We call

$$\mathrm{R}\Delta\, S(\boldsymbol{x}, z) = \mathsf{PU}(\hat{\theta}, z) = \Sigma\Delta\, S(\boldsymbol{x}, z) - \mathrm{C}\Delta\, S(\boldsymbol{x}, z) \tag{30}$$

the *pseudo-residual* since it is a weighted residual in the sense of [6].

The pseudo-residual serves as a surrogate for the numerator in the score test statistic (26). For the denominator, we need the variance of the pseudo-residual. Appendix B gives an exact formula (65) for the variance of the pseudo-score $\mathsf{PU}(\theta, z)$, which can serve as an approximation to the variance of the pseudo-residual $\mathrm{R}\Delta\, S(\boldsymbol{x}, z)$. The leading term in this approximation is

$$\mathrm{C}^2\Delta\, S(\boldsymbol{x}, z) = \int_W [\Delta_u S(\boldsymbol{x}, z)]^2 \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u \tag{31}$$

which we shall call the *Poincaré pseudo-variance* because of its similarity to the Poincaré upper bound in (21). We propose to use the square root of (31) as a surrogate for the denominator in (26). This yields a *'standardised' pseudo-residual*

$$\mathrm{T}\Delta\, S(\boldsymbol{x}, z) = \mathrm{R}\Delta\, S(\boldsymbol{x}, z) / \sqrt{\mathrm{C}^2\Delta\, S(\boldsymbol{x}, z)}. \tag{32}$$

We emphasise that this quantity is not guaranteed to have zero mean and unit variance (even approximately) under the null hypothesis. It is a computationally efficient surrogate for the score test statistic; its null distribution must be investigated by other means.

The pseudo-sum (28) can be regarded as a functional summary statistic for the data in its own right. Its definition depends only on the choice of the statistic S, and it may have a meaningful interpretation as a non-parametric estimator of a property of the point process. The pseudo-compensator (29) might also be regarded as a functional summary statistic, but its definition involves the null model. If the null model is true we may expect the pseudo-residual to be approximately zero. Sections 9-11 and Appendix A study particular instances of pseudo-residual diagnostics based on (28)-(30).

In the conditional case, the Papangelou conditional intensity $\lambda_{\hat{\theta}}(u, \boldsymbol{x})$ must be replaced by $\lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)$ given in (19) or (24). The integral in the definition of the pseudo-compensator (29) must be restricted to the domain W°, and the summation over data points in (28) must be restricted to points $x_i \in W^\circ$, i.e. to summation over points of $\boldsymbol{x}^\circ$.

6.3 Residuals

A simpler surrogate for the score test is available when the canonical sufficient statistic S of the perturbing model is naturally expressible as a sum of local contributions

$$S(\boldsymbol{x}, z) = \sum_i s(x_i, \boldsymbol{x}_{-i}, z). \tag{33}$$

Note that any statistic can be decomposed in this way unless some restriction is imposed on s; such a decomposition is not necessarily unique. We call the decomposition 'natural' if $s(u, \boldsymbol{x}, z)$ only depends on points of $\boldsymbol{x}$ that are close to u, as demonstrated in the examples in Sections 9, 10 and 11 and in Appendix A.

Consider a null model with conditional intensity $\lambda_\theta(u, \boldsymbol{x})$. Following [6] define the (s-weighted) innovation by

$$\mathrm{I}\, S(\boldsymbol{x}, r) = S(\boldsymbol{x}, z) - \int_W s(u, \boldsymbol{x}, z) \lambda_\theta(u, \boldsymbol{x}) \, \mathrm{d}u \tag{34}$$

which by the GNZ formula (17) has mean zero under the null model. In practice we replace θ by an estimate $\hat{\theta}$ (e.g. the MPLE) and consider the *(s-weighted) residual*

$$\mathrm{R}\, S(\boldsymbol{x}, z) = S(\boldsymbol{x}, z) - \int_W s(u, \boldsymbol{x}, z) \lambda_{\hat{\theta}}(u, \boldsymbol{x}) \, \mathrm{d}u. \tag{35}$$

The residual shares many properties of the score function and can serve as a computationally efficient surrogate for the score. The data-dependent integral

$$\mathrm{C}\, S(\boldsymbol{x}, z) = \int_W s(u, \boldsymbol{x}, z) \lambda_{\hat{\theta}}(u, \boldsymbol{x}) \, \mathrm{d}u \tag{36}$$

is the *(estimated) Papangelou compensator* of S. By the general variance formula (64) and by analogy with (31) we propose to use the *Poincaré variance*

$$\mathrm{C}^2\, S(\boldsymbol{x}, z) = \int_W s(u, \boldsymbol{x}, z)^2 \lambda_{\hat{\theta}}(u, \boldsymbol{x}) \, \mathrm{d}u \tag{37}$$

as a surrogate for the variance of $\mathrm{R}\, S(\boldsymbol{x}, z)$, and thereby obtain a 'standardised' residual

$$\mathrm{T}\, S(\boldsymbol{x}, z) = \mathrm{R}\, S(\boldsymbol{x}, z) / \sqrt{\mathrm{C}^2\, S(\boldsymbol{x}, z)}.$$

Once again $\mathrm{T}\, S(\boldsymbol{x}, z)$ is not exactly standardised, and its null distribution must be investigated by other means.

In the conditional case, the integral in the definition of the compensator (36) must be restricted to the domain W°, and the summation over data points in (33) must be restricted to points $x_i \in W^\circ$, i.e. to summation over points of $\boldsymbol{x}^\circ$.

7. DIAGNOSTICS FOR FIRST ORDER TREND

Consider any null model with density $f_\theta(\boldsymbol{x})$ and conditional intensity $\lambda_\theta(u, \boldsymbol{x})$. By analogy with Section 4 we consider alternatives of the form (25) where

$$S(\boldsymbol{x}, z) = \sum_i s(x_i, z)$$

for some function s. The perturbing model (27) is a Poisson process with intensity $\exp(\phi s(\cdot, z))$ where z is a nuisance parameter. The score test is a test for the presence of an (extra) first order trend. The pseudo-score and residual diagnostics are both equal to

$$\mathrm{R}\, S(\boldsymbol{x}, z) = \sum_i s(x_i, z) - \int_W s(u, z)\lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u. \tag{38}$$

This is the s-weighted residual described in [6]. The variance of (38) can be estimated by simulation, or approximated by the Poincaré variance (37).

If Z is a real-valued covariate function on W then we may take $s(u, z) = \mathbb{I}\{Z(u) \le z\}$ for $z \in \mathbb{R}$, corresponding to a threshold effect (cf. Section 4.2). A plot of (38) against z was called a *lurking variable plot* in [6].

If $s(u, z) = k(u - z)$ for $z \in \mathbb{R}^2$, where k is a density function on $\mathbb{R}^2$, then

$$\mathrm{R}\, S(\boldsymbol{x}, z) = \sum_i k(x_i - z) - \int_W k(u - z)\lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u$$

which was dubbed the *smoothed residual field* in [6]. Examples of application of these techniques have been discussed extensively in [6].

8. INTERPOINT INTERACTION

In the remainder of the paper we concentrate on diagnostics for interpoint interaction.

8.1 Classical summary statistics

Following Ripley's influential paper [63] it is standard practice, when investigating association or dependence between points in a spatial point pattern, to evaluate functional summary statistics such as the K-function, and to compare graphically the empirical summaries and theoretical predicted values under a suitable model, often a stationary Poisson process ('Complete Spatial Randomness', CSR) [63, 22, 28].

The three most popular functional summary statistics for spatial point processes are Ripley's K-function, the nearest neighbour distance distribution function G, and the empty space function (spherical contact distance distribution function) F. Definitions of K, G and F and their estimators can be seen in [9, 22, 28, 50]. Simple empirical estimators of these functions are of the form

$$\hat{K}(r) = \hat{K}_{\boldsymbol{x}}(r) = \frac{1}{\widehat{\rho^2}(\boldsymbol{x})|W|} \sum_{i \ne j} e_K(x_i, x_j)\mathbb{I}\{\|x_i - x_j\| \le r\} \tag{39}$$

$$\hat{G}(r) = \hat{G}_{\boldsymbol{x}}(r) = \frac{1}{n(\boldsymbol{x})} \sum_i e_G(x_i, \boldsymbol{x}_{-i}, r)\mathbb{I}\{d(x_i, \boldsymbol{x}_{-i}) \le r\} \tag{40}$$

$$\hat{F}(r) = \hat{F}_{\boldsymbol{x}}(r) = \frac{1}{|W|} \int_W e_F(u, r)\mathbb{I}\{d(u, \boldsymbol{x}) \le r\}\, \mathrm{d}u \tag{41}$$

where $e_K(u, v)$, $e_G(u, \boldsymbol{x}, r)$ and $e_F(u, r)$ are edge correction weights, and typically $\widehat{\rho^2}(\boldsymbol{x}) = n(\boldsymbol{x})(n(\boldsymbol{x}) - 1)/|W|^2$.

8.2 Score test approach

The classical approach fits naturally into the scheme of Section 6. In order to test for dependence between points, we choose a perturbing model that exhibits dependence. Three interesting examples of perturbing models are the Strauss process, the Geyer saturation model with saturation threshold 1 and the area-interaction process, with interaction potentials $V_S(\boldsymbol{x}, r)$, $V_G(\boldsymbol{x}, r)$ and $V_A(\boldsymbol{x}, r)$ given in (14)-(16). The nuisance parameter $r \geq 0$ determines the range of interaction. Although the Strauss density is integrable only when $\phi \leq 0$, a Strauss hybrid (between f_θ and the Strauss density) may be well-defined for some $\phi > 0$ so that the extended model may support alternatives that are clustered relative to the null, as originally intended by Strauss [71].

The potentials of these three models are closely related to the summary statistics $\hat{K}, \hat{G}$ and $\hat{F}$ in (39)–(41). Ignoring the edge correction weights $e(\cdot)$ we have

$$\hat{K}_{\boldsymbol{x}}(r) \approx \frac{2|W|}{n(\boldsymbol{x})\,(n(\boldsymbol{x})-1)} V_S(\boldsymbol{x}, r) \tag{42}$$

$$\hat{G}_{\boldsymbol{x}}(r) \approx \frac{1}{n(\boldsymbol{x})} V_G(\boldsymbol{x}, r) \tag{43}$$

$$\hat{F}_{\boldsymbol{x}}(r) \approx -\frac{1}{|W|} V_A(\boldsymbol{x}, r). \tag{44}$$

To draw the closest possible connection with the score test, instead of choosing the Strauss, Geyer or area-interaction process as the perturbing model, we shall take the perturbing model to be defined through (27) where S is one of the statistics $\hat{K}$, $\hat{G}$ or $\hat{F}$. We call these the *(perturbing) $\hat{K}$-model*, *$\hat{G}$-model* and *$\hat{F}$-model* respectively. The score test is then precisely equivalent to comparing $\hat{K}$, $\hat{G}$ or $\hat{F}$ with its predicted expectation using (6).

Essentially $\hat{K}$, $\hat{G}$, $\hat{F}$ are re-normalised versions of V_S, V_G, V_A as shown in (42)–(44). In the case of $\hat{F}$ the renormalisation is not data-dependent, so the $\hat{F}$-model is virtually an area-interaction model, ignoring edge correction. For $\hat{K}$, the renormalisation depends only on $n(\boldsymbol{x})$, and so conditionally on $n(\boldsymbol{x}) = n$, the $\hat{K}$-model and the Strauss process are approximately equivalent. Similarly for $\hat{G}$, the normalisation also depends only on $n(\boldsymbol{x})$, so conditionally on $n(\boldsymbol{x}) = n$, the $\hat{G}$-model and Geyer saturation process are approximately equivalent. If we follow Ripley's [63] recommendation to condition on n when testing for interaction, this implies that the use of the K, G or F-function is approximately equivalent to the score test of CSR against a Strauss, Geyer or area-interaction alternative, respectively.

When the null hypothesis is CSR, we saw that the extended model (25) is identical to the perturbing model, up to a change in intensity, so that the use of the $\hat{K}$-function is equivalent to testing the null hypothesis of CSR against the alternative of a $\hat{K}$-model. Similarly for $\hat{G}$ and $\hat{F}$. For a more general null hypothesis, the use of the $\hat{K}$-function, for example, corresponds to adopting an alternative hypothesis that is a hybrid between the fitted model and a $\hat{K}$-model.

Note that if the edge correction weight $e_K(u, v)$ is uniformly bounded, the $\hat{K}$-model is integrable for all values of ϕ, avoiding a difficulty with the Strauss process [42].

Computation of the score test statistic (26) requires estimation or approximation of the null variance of $\hat{K}(r)$, $\hat{G}(r)$ or $\hat{F}(r)$. A wide variety of approximations

is available when the null hypothesis is CSR [64, 28]. For other null hypotheses, simulation estimates would typically be used. A central limit theorem is available for $\hat{K}(r)$, $\hat{G}(r)$ and $\hat{F}(r)$, e.g. [7, 34, 33, 39, 64]. However, convergence is not uniform in r, and the normal approximation will be poor for small values of r. Instead Ripley [62] developed an exact Monte Carlo test [12, 35] based on simulation envelopes of the summary statistic under the null hypothesis.

In the following sections we develop the residual and pseudo-residual diagnostics corresponding to this approach.

9. RESIDUAL DIAGNOSTICS FOR INTERACTION USING PAIRWISE DISTANCES

This section develops residual (35) and pseudo-residual (30) diagnostics derived from a summary statistic S which is a sum of contributions depending on pairwise distances.

9.1 Residual based on perturbing Strauss model

9.1.1 General derivation Consider any statistic of the general 'pairwise interaction' form

$$S(\boldsymbol{x}, r) = \sum_{i<j} q(\{x_i, x_j\}, r). \tag{45}$$

This can be decomposed in the local form (33) with

$$s(u, \boldsymbol{x}, r) = \frac{1}{2} \sum_i q(\{x_i, u\}, r), \quad u \notin \boldsymbol{x}.$$

Hence

$$\Delta_{x_i} S(\boldsymbol{x}, r) = 2s(x_i, \boldsymbol{x}_{-i}, r) \quad \text{and} \quad \Delta_u S(\boldsymbol{x}, r) = 2s(u, \boldsymbol{x}, r), \quad u \notin \boldsymbol{x}.$$

Consequently the pseudo-residual and the pseudo-compensator are just twice the residual and the Papangelou compensator:

$$\Sigma\Delta\, S(\boldsymbol{x}, r) \;=\; 2S(\boldsymbol{x}, r) \;= \sum_{i \neq j} q(\{x_i, x_j\}, r) \tag{46}$$

$$\mathrm{C}\Delta\, S(\boldsymbol{x}, r) \;=\; 2\,\mathrm{C}\, S(\boldsymbol{x}, r) = \int_W \sum_i q(\{x_i, u\}, r) \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u \tag{47}$$

$$\mathrm{R}\Delta\, S(\boldsymbol{x}, z) \;=\; 2\,\mathrm{R}\, S(\boldsymbol{x}, r) = 2S(\boldsymbol{x}, r) - 2\,\mathrm{C}\, S(\boldsymbol{x}, r). \tag{48}$$

9.1.2 Residual of Strauss potential The Strauss interaction potential V_S of (14) is of the general form (45) with $q(\{x_i, x_j\}, r) = \mathbb{I}\{\|x_i - x_j\| \leq r\}$. Hence V_S can be decomposed in the form (33) with $s(u, \boldsymbol{x}, r) = \frac{1}{2} t(u, \boldsymbol{x}, r)$ where

$$t(u, \boldsymbol{x}, r) = \sum_i \mathbb{I}\{\|u - x_i\| \leq r\}, \quad u \notin \boldsymbol{x}.$$

Hence the Papangelou compensator of V_S is

$$\mathrm{C}\, V_S(\boldsymbol{x}, r) = \frac{1}{2} \int_W t(u, \boldsymbol{x}, r) \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u. \tag{49}$$

9.1.3 Case of CSR If the null model is CSR with intensity ρ estimated by $\hat{\rho} = n(\boldsymbol{x})/|W|$ (the MLE, which agrees with the MPLE in this case), the Papangelou compensator (49) becomes

$$\mathrm{C}\, V_S(\boldsymbol{x}, r) = \frac{\hat{\rho}}{2} \int_W \sum_i \mathbb{I}\{\|u - x_i\| \le r\} \,\mathrm{d}u = \frac{\hat{\rho}}{2} \sum_i |W \cap B(x_i, r)|.$$

Ignoring edge effects we have $|W \cap B(x_i, r)| \approx \pi r^2$ and, applying (42), the residual is approximately

$$\mathrm{R}\, V_S(\boldsymbol{x}, r) \approx \frac{n(\boldsymbol{x})^2}{2|W|} \left[\hat{K}_{\boldsymbol{x}}(r) - \pi r^2\right]. \tag{50}$$

The term in brackets is a commonly-used measure of departure from CSR, and is a sensible diagnostic because $K(r) = \pi r^2$ under CSR. The Poincaré variance (37) is

$$\mathrm{C}^2\, V_S(\boldsymbol{x}, r) = \frac{n(\boldsymbol{x})}{4|W|} \int_W t(u, \boldsymbol{x}, r)^2 \,\mathrm{d}u$$

while the exact variance formula (64) yields

$$\begin{aligned} \mathbb{V}\mathrm{ar}\,[\mathrm{R}\, V_S(\boldsymbol{X}, r)] &\approx \mathbb{V}\mathrm{ar}\,[\mathrm{I}\, V_S(\boldsymbol{X}, r)] \\ &= \frac{\rho}{4} \int_W \mathbb{E}\left[t(u, \boldsymbol{X}, r)^2\right] \,\mathrm{d}u + \frac{\rho^2}{4} \int_{W^2} \mathbb{I}\{\|u - v\| \le r\} \,\mathrm{d}u \,\mathrm{d}v. \end{aligned}$$

Now $Y = t(u, \boldsymbol{X}, r)$ is Poisson distributed with mean $\mu = \rho |B(u, r) \cap W|$ so that $\mathbb{E}(Y^2) = \mu + \mu^2$. For $u \in W_{\ominus r}$ we have $\mu = \rho\pi r^2$, so ignoring edge effects

$$\mathbb{V}\mathrm{ar}\,[\mathrm{R}\, V_S(\boldsymbol{X}, r)] \approx \frac{\rho^2}{2} |W| \pi r^2 + \frac{\rho^3}{4} |W| \pi^2 r^4.$$

This has similar functional form to expressions for the variance of $\hat{K}$ under CSR obtained using the methods of U-statistics [47, 16, 64], summarised in [28, p. 51 ff.]. For small r, we have $t(u, \boldsymbol{x}, r) \in \{0, 1\}$ so that

$$\begin{aligned} \mathrm{C}^2\, V_S(\boldsymbol{x}, r) &\approx \frac{n(\boldsymbol{x})^2}{4|W|} \pi r^2 \\ \mathbb{V}\mathrm{ar}\,[\mathrm{R}\, V_S(\boldsymbol{X}, r)] &\approx \frac{\rho^2}{2} |W| \pi r^2 \end{aligned}$$

so that $\mathrm{C}^2\, V_S(\boldsymbol{x}, r)$ is a substantial underestimate (by a factor of approximately 2) of the true variance. Thus a test based on referring $\mathrm{T}\, V_S(\boldsymbol{x}, r)$ to a standard normal distribution may be expected to be conservative for small r.

9.2 Residual based on perturbing $\hat{K}$-model

Assuming $\widehat{\rho^2}(\boldsymbol{x}) = \widehat{\rho^2}(n(\boldsymbol{x}))$ depends only on $n(\boldsymbol{x})$, the empirical K-function (39) can also be expressed as a sum of local contributions $\hat{K}_{\boldsymbol{x}}(r) = \sum_i k(x_i, \boldsymbol{x}_{-i}, r)$ with

$$k(u, \boldsymbol{x}, r) = \frac{t^w(u, \boldsymbol{x}, r)}{\widehat{\rho^2}(n(\boldsymbol{x}) + 1)|W|}, \quad u \notin \boldsymbol{x}$$

where

$$t^w(u, \boldsymbol{x}, r) = \sum_j e_K(u, x_j)\mathbb{I}\{\|u - x_j\| \le r\}$$

is a weighted count of the points of $\boldsymbol{x}$ that are r-close to the location u. Hence the compensator of the $\hat{K}$-function is

$$\mathrm{C}\,\hat{K}_{\boldsymbol{x}}(r) = \frac{1}{\widehat{\rho^2}(n(\boldsymbol{x}) + 1)|W|} \int_W t^w(u, \boldsymbol{x}, r)\lambda_{\hat{\theta}}(u, \boldsymbol{x})\,\mathrm{d}u. \tag{51}$$

Assume the edge correction weight $e_K(u, v) = e_K(v, u)$ is symmetric; e.g. this is satisfied by the Ohser-Stoyan edge correction weight [57, 56] given by $e_K(u, v) = 1/|W_u \cap W_v|$ where $W_u = \{u + v : v \in W\}$, but not by Ripley's [62] isotropic correction weight. Then the increment is, for $u \notin \boldsymbol{x}$,

$$\Delta_u \hat{K}_{\boldsymbol{x}}(r) = \frac{\widehat{\rho^2}(\boldsymbol{x}) - \widehat{\rho^2}(\boldsymbol{x} \cup \{u\})}{\widehat{\rho^2}(\boldsymbol{x} \cup \{u\})}\hat{K}_{\boldsymbol{x}}(r) + \frac{2t^w(u, \boldsymbol{x}, r)}{\widehat{\rho^2}(\boldsymbol{x} \cup \{u\})|W|}$$

and when $x_i \in \boldsymbol{x}$

$$\Delta_{x_i} \hat{K}_{\boldsymbol{x}}(r) = \frac{\widehat{\rho^2}(\boldsymbol{x}_{-i}) - \widehat{\rho^2}(\boldsymbol{x})}{\widehat{\rho^2}(\boldsymbol{x}_{-i})}\hat{K}_{\boldsymbol{x}}(r) + \frac{2t^w(x_i, \boldsymbol{x}_{-i}, r)}{\widehat{\rho^2}(\boldsymbol{x}_{-i})|W|}.$$

Assuming the standard estimator $\widehat{\rho^2}(\boldsymbol{x}) = n(n-1)/|W|^2$ with $n = n(\boldsymbol{x})$, the pseudo-sum is seen to be zero, so the pseudo-residual is apart from the sign equal to the pseudo-compensator, which becomes

$$\mathrm{C}\Delta\,\hat{K}_{\boldsymbol{x}}(r) = 2\,\mathrm{C}\,\hat{K}_{\boldsymbol{x}}(r) - \left[\frac{2}{n-2}\int_W \lambda_{\hat{\theta}}(u, \boldsymbol{x})\,\mathrm{d}u\right]\hat{K}_{\boldsymbol{x}}(r)$$

where $\mathrm{C}\,\hat{K}_{\boldsymbol{x}}(r)$ is given by (51). So if the null model is CSR and the intensity is estimated by $n/|W|$, the pseudo-residual is approximately $2[\hat{K}_{\boldsymbol{x}}(r) - \mathrm{C}\,\hat{K}_{\boldsymbol{x}}(r)]$, and hence it is equivalent to the residual approximated by (50). This is also the conclusion in the more general case of a null model with an activity parameter κ, i.e. where the conditional intensity factorises as

$$\lambda_\theta(u, \boldsymbol{x}) = \kappa\xi_\beta(u, \boldsymbol{x})$$

where $\theta = (\kappa, \beta)$ and $\xi_\beta(\cdot)$ is a conditional intensity, since the pseudo-likelihood equations then imply that $n = \int_W \lambda_{\hat{\theta}}(u, \boldsymbol{x})\,\mathrm{d}u$.

In conclusion, the residual diagnostics obtained from the perturbing Strauss and $\hat{K}$-models are very similar, the major difference being the data-dependent normalisation of the $\hat{K}$-function; similarly for pseudo-residual diagnostics which may be effectively equivalent to the residual diagnostics. In practice, the popularity of the K-function seems to justify using the residual diagnostics based on the perturbing $\hat{K}$-model. Furthermore, due to the familarity of the K-function we often choose to plot the compensator(s) of the fitted model(s) in a plot with the empirical K-function rather than the residual(s) for the fitted model.

9.3 Edge correction in conditional case

In the conditional case, the conditional intensity $\lambda_{\hat{\theta}}(u, \boldsymbol{x})$ is known only at locations $u \in W^\circ$. The diagnostics must be modified accordingly, by restricting the domain of summation and integration to W°. Appropriate modifications are discussed in Appendices C–E.

10. RESIDUAL DIAGNOSTICS FOR INTERACTION USING NEAREST NEIGHBOUR DISTANCES

This section develops residual and pseudo-residual diagnostics derived from summary statistics based on nearest neighbour distances.

10.1 Residual based on perturbing Geyer model

The Geyer interaction potential $V_G(\boldsymbol{x}, r)$ given by (15) is clearly a sum of local statistics (33), and its compensator is

$$\mathrm{C}\, V_G(\boldsymbol{x}, r) = \int_W \mathbb{I}\{d(u, \boldsymbol{x}) \le r\} \lambda_{\hat\theta}(u, \boldsymbol{x})\, \mathrm{d}u.$$

The Poincaré variance is equal to the compensator in this case. Ignoring edge effects, $V_G(\boldsymbol{x}, r)$ is approximately $n(\boldsymbol{x})\hat{G}_{\boldsymbol{x}}(r)$, cf. (40).

If the null model is CSR with estimated intensity $\hat\kappa = n(\boldsymbol{x})/|W|$, then

$$\mathrm{C}\, V_G(\boldsymbol{x}, r) = \hat\kappa |W \cap \bigcup_i B(x_i, r)|;$$

ignoring edge effects, this is approximately $\hat\kappa |W| \hat{F}(r)$, cf. (41). Thus the residual diagnostic is approximately $n(\boldsymbol{x})(\hat{G}(r) - \hat{F}(r))$. This is a reasonable diagnostic for departure from CSR, since $F \equiv G$ under CSR. This argument lends support to Diggle's [26, eq. (5.7)] proposal to judge departure from CSR using the quantity $\sup |\hat{G} - \hat{F}|$.

This example illustrates the important point that the compensator of a functional summary statistic S should not be regarded as an alternative parametric estimator of the same quantity that S is intended to estimate. In the example just given, under CSR the compensator of $\hat{G}$ is approximately $\hat{F}$, a qualitatively different and in some sense 'opposite' summary of the point pattern.

We have observed that the interaction potential V_G of the Geyer saturation model is closely related to $\hat{G}$. However, the pseudo-residual associated to V_G is a more complicated statistic, since a straightforward calculation shows that the pseudo-sum is

$$\Sigma\Delta\, V_G(\boldsymbol{x}, r) = V_G(\boldsymbol{x}, r) + \sum_i \sum_{j:\, j \ne i} \mathbb{I}\{\|x_i - x_j\| \le r \text{ and } d(x_j, \boldsymbol{x}_{-i}) > r\},$$

and the pseudo-compensator is

$$\begin{aligned} \mathrm{C}\Delta\, V_G(\boldsymbol{x}, r) = &\int_W \mathbb{I}\{d(u, \boldsymbol{x}) \le r\} \lambda_{\hat\theta}(u, \boldsymbol{x})\, \mathrm{d}u \\ &+ \sum_i \mathbb{I}\{d(x_i, \boldsymbol{x}_{-i}) > r\} \int_W \mathbb{I}\{\|u - x_i\| \le r\} \lambda_{\hat\theta}(u, \boldsymbol{x})\, \mathrm{d}u. \end{aligned}$$

10.2 Residual based on perturbing $\hat{G}$-model

The empirical G-function (40) can be written

$$\hat{G}_{\boldsymbol{x}}(r) = \sum_i g(x_i, \boldsymbol{x}_{-i}, r) \tag{52}$$

where

$$g(u, \boldsymbol{x}, r) = \frac{1}{n(\boldsymbol{x}) + 1} e_G(u, \boldsymbol{x}, r) \mathbb{I}\{d(u, \boldsymbol{x}) \leq r\}, \quad u \notin \boldsymbol{x} \tag{53}$$

so that the Papangelou compensator of the empirical G-function is

$$\mathrm{C}\, \hat{G}_{\boldsymbol{x}}(r) = \int_W g(u, \boldsymbol{x}, r) \lambda_{\hat{\theta}}(u, \boldsymbol{x}) \, \mathrm{d}u = \frac{1}{n(\boldsymbol{x}) + 1} \int_{W \cap \bigcup_i B(x_i, r)} e_G(u, \boldsymbol{x}, r) \lambda_{\hat{\theta}}(u, \boldsymbol{x}) \, \mathrm{d}u.$$

The residual diagnostics obtained from the Geyer and $\hat{G}$-models are very similar, and we choose to use the diagnostic based on the popular $\hat{G}$-function. As with the K-function we typically use the compensator(s) of the fitted model(s) rather than the residual(s), to visually maintain the close connection to the empirical G-function.

The expressions for the pseudo-sum and pseudo-compensator of $\hat{G}$ are not of simple form, and we refrain from explicitly writing out these expressions. For both the $\hat{G}$- and Geyer models, the pseudo-sum and pseudo-compensator are not directly related to a well-known summary statistic. We prefer to plot the pseudo-residual rather than the pseudo-sum and pseudo-compensator(s).

11. DIAGNOSTICS FOR INTERACTION BASED ON EMPTY SPACE DISTANCES

11.1 Pseudo-residual based on perturbing area-interaction model

When the perturbing model is the area-interaction process, it is convenient to re-parametrise the density, such that the canonical sufficient statistic V_A given in (16) is re-defined as

$$V_A(\boldsymbol{x}, r) = \frac{1}{|W|} |W \cap \bigcup_i B(x_i, r)|.$$

This summary statistic is not naturally expressed as a sum of contributions from each point as in (33), so we shall only construct the pseudo-residual. Let

$$U(\boldsymbol{x}, r) = W \cap \bigcup_i B(x_i, r).$$

The increment

$$\Delta_u V_A(\boldsymbol{x}, r) = \frac{1}{|W|} \left(|U(\boldsymbol{x} \cup \{u\}, r)| - |U(\boldsymbol{x}, r)| \right), \quad u \notin \boldsymbol{x}$$

can be thought of as 'unclaimed space' — the proportion of space around the location u that is not "claimed" by the points of $\boldsymbol{x}$. The pseudo-sum

$$\Sigma\Delta\, V_A(\boldsymbol{x}, r) = \sum_i \Delta_{x_i} V_A(\boldsymbol{x}, r)$$

is the proportion of the window that has 'single coverage' — the proportion of locations in W that are covered by exactly one of the balls $B(x_i, r)$. This can be used in its own right as a functional summary statistic, and it corresponds to

a raw (i.e. not edge corrected) empirical estimate of a summary function $F_1(r)$ defined by

$$F_1(r) = \mathbb{P}\left(\#\{x \in \boldsymbol{X} | d(u,x) \le r\} = 1\right),$$

for any stationary point process $\boldsymbol{X}$, where $u \in \mathbb{R}^2$ is arbitrary. Under CSR with intensity ρ we have

$$\mathbb{E}F_1(r) = \rho\pi r^2 \exp(-\rho\pi r^2).$$

This summary statistic does not appear to be treated in the literature, and it may be of interest to study it separately, but we refrain from a more detailed study here.

The pseudo-compensator corresponding to this pseudo-sum is

$$\mathrm{C}\Delta\, V_A(\boldsymbol{x}, r) = \int_W \Delta_u V_A(\boldsymbol{x}, r) \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u.$$

This integral does not have a particularly simple interpretation even when the null model is CSR.

11.2 Pseudo-residual based on perturbing $\hat{F}$-model

Alternatively one could use a standard empirical estimator $\hat{F}$ of the empty space function F as the summary statistic in the pseudo-residual. The pseudo-sum associated with the perturbing $\hat{F}$-model is

$$\Sigma\Delta\, \hat{F}_{\boldsymbol{x}}(r) = n(\boldsymbol{x})\hat{F}_{\boldsymbol{x}}(r) - \sum_i \hat{F}_{\boldsymbol{x}_{-i}}(r),$$

with pseudo-compensator

$$\mathrm{C}\Delta\, \hat{F}_{\boldsymbol{x}}(r) = \int_W (\hat{F}_{\boldsymbol{x}\cup\{u\}}(r) - \hat{F}_{\boldsymbol{x}}(r)) \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u.$$

Ignoring edge correction weights, $\hat{F}_{\boldsymbol{x}\cup\{u\}}(r) - \hat{F}_{\boldsymbol{x}}(r)$ is approximately equal to $\Delta_u V_A(\boldsymbol{x}, r)$, so the pseudo-sum and pseudo-compensator associated with the perturbing $\hat{F}$-model are approximately equal to the pseudo-sum and pseudo-compensator associated with the perturbing area-interaction model. Here, we usually prefer graphics using the pseudo-compensator(s) and the pseudo-sum since this has an intuitive interpretation as explained above.

12. TEST CASE: TREND WITH INHIBITION

In Sections 12–14 we demonstrate the diagnostics on the point pattern datasets shown in Figure 1. This section concerns the synthetic point pattern in Figure 1b.

12.1 Data and models

Figure 1b shows a simulated realisation of the inhomogeneous Strauss process with first order term $\lambda(x, y) = 200\exp(2x + 2y + 3x^2)$, interaction range $R = 0.05$, interaction parameter $\gamma = \exp(\phi) = 0.1$ and W equal to the unit square, see (13) and (14). This is an example of extremely strong inhibition (negative association) between neighbouring points, combined with a spatial trend. Since it is easy to recognise spatial trend in the data, (either visually or using existing tools such as kernel smoothing [27]) the main challenge here is to detect the inhibition after accounting for the trend.

We fitted four point process models to the data in Figure 1b. They were (A) a homogeneous Poisson process (CSR); (B) an inhomogeneous Poisson process with the correct form of the first order term, i.e. with intensity

$$\rho(x,y) = \exp(\beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2) \tag{54}$$

where $\beta_0, \ldots, \beta_3$ are real parameters; (C) a homogeneous Strauss process with the correct interaction range $R = 0.05$; and (D) a process of the correct form, i.e. inhomogeneous Strauss with the correct interaction range $R = 0.05$ and the correct form of the first order potential (54).

12.2 Software implementation

The diagnostics defined in Sections 9–11 were implemented in the R language, and will be publicly available in the `spatstat` library [5]. Unless otherwise stated, models were fitted by approximate maximum pseudo-likelihood using the algorithm of [4] with the default quadrature scheme in `spatstat`, having an $m \times m$ grid of dummy points where $m = \max(25, 10[1 + 2\sqrt{n(\boldsymbol{x})}/10])$ was equal to 40 for most of our examples. Integrals over the domain W were approximated by finite sums over the quadrature points.

Some models were refitted using a finer grid of dummy points, usually 80×80. The software also supports Huang-Ogata [36] one-step approximate maximum likelihood.

12.3 Application of $\hat{K}$ diagnostics

12.3.1 Diagnostics for correct model First we fitted a point process model of the correct form (D). The fitted parameter values were $\hat{\gamma} = 0.217$ and $\hat{\beta} = (5.6, -0.46, 3.35, 2.05)$ using the coarse grid of dummy points, and $\hat{\gamma} = 0.170$ and $\hat{\beta} = (5.6, -0.64, 4.06, 2.44)$ using the finer grid of dummy points, as against the true values $\gamma = 0.1$ and $\beta = (5.29, 2, 2, 3)$.

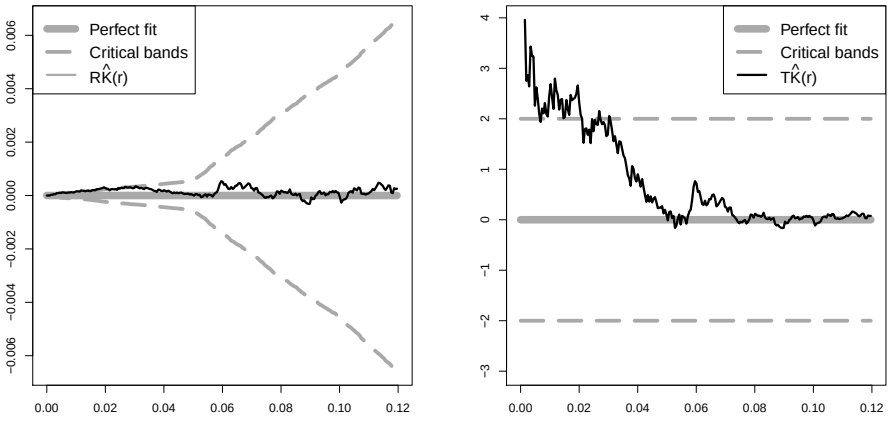


Fig 3: Residual diagnostics based on pairwise distances, for a model of the correct form fitted to the data in Figure 1b. (a) residual $\hat{K}$-function and two-standard-deviation limits under the fitted model of the correct form. (b) standardised residual $\hat{K}$-function under the fitted model of the correct form.

Figure 2 in Section 1 shows $\hat{K}$ along with its compensator for the fitted model, together with the theoretical K-function under CSR. The empirical K-function and its compensator coincide very closely, suggesting correctly that the model is a good fit. Figure 3a shows the residual $\hat{K}$-function and the two-standard-deviation limits, where the surrogate standard deviation is the square root of (37). Figure 3b shows the corresponding standardised residual $\hat{K}$-function obtained by dividing by the surrogate standard deviation.

Although this model is of the correct form, the standardised residual exceeds 2 for small values of r. This is consistent with the prediction in Section 9.1.3 that the test would be conservative for small r. For very small r there are small-sample effects so that a normal approximation to the null distribution of the standardised residual is inappropriate.

Formal significance interpretation of the critical bands is limited, because the null distribution of the standardised residual is not known exactly, and the values ± 2 are approximate *pointwise* critical values, i.e. critical values for the score test based on fixed r. The usual problems of multiple testing arise when the test statistic is considered as a function of r: see [28, p. 14].

12.3.2 Comparison of competing models Figure 4a shows the empirical K-function and its compensator for each of the models (A)–(D) in Section 12.1. Figure 4b shows the corresponding residual plots, and Figure 4c the standardised residuals. A positive or negative value of the residual suggests that the data are more clustered or more inhibited, respectively, than the model. The clear inference is that the Poisson models (A) and (B) fail to capture interpoint inhibition at range $r \approx 0.05$, while the homogeneous Strauss model (C) is less clustered than the data at very large scales, suggesting that it fails to capture spatial trend. The correct model (D) is judged to be a good fit.

The interpretation of this example requires some caution, because the residual $\hat{K}$-function of the fitted Strauss models (C) and (D) is constrained to be approximately zero at $r = R = 0.05$. The maximum pseudo-likelihood fitting algorithm solves an estimating equation that is approximately equivalent to this constraint, because of (42).

It is debatable which of the presentations in Figure 4 is more effective at revealing lack-of-fit. A compensator plot such as Figure 4a seems best at capturing the main differences between competing models. It is particularly useful for recognising a gross lack-of-fit. A residual plot such as Figure 4b seems better for making finer comparisons of goodness-of-fit, for example, assessing models with slightly different ranges of interaction. A standardised residual plot such as Figure 4c tends to be highly irregular for small values of r, due to discretisation effects in the computation and the inherent nondifferentiability of the empirical statistic. In difficult cases we may apply smoothing to the standardised residual.

12.4 Application of $\hat{G}$ diagnostics

12.4.1 Diagnostics for correct model Consider again the model of the correct form (D). The residual and compensator of the empirical nearest neighbour function $\hat{G}$ for the fitted model are shown in Figure 5. The residual plot suggests a marginal lack-of-fit for $r < 0.025$. This may be correct, since the fitted model parameters (Section 12.3.1) are marginally poor estimates of the true values, in particular of the interaction parameter. This was not reflected so strongly in the

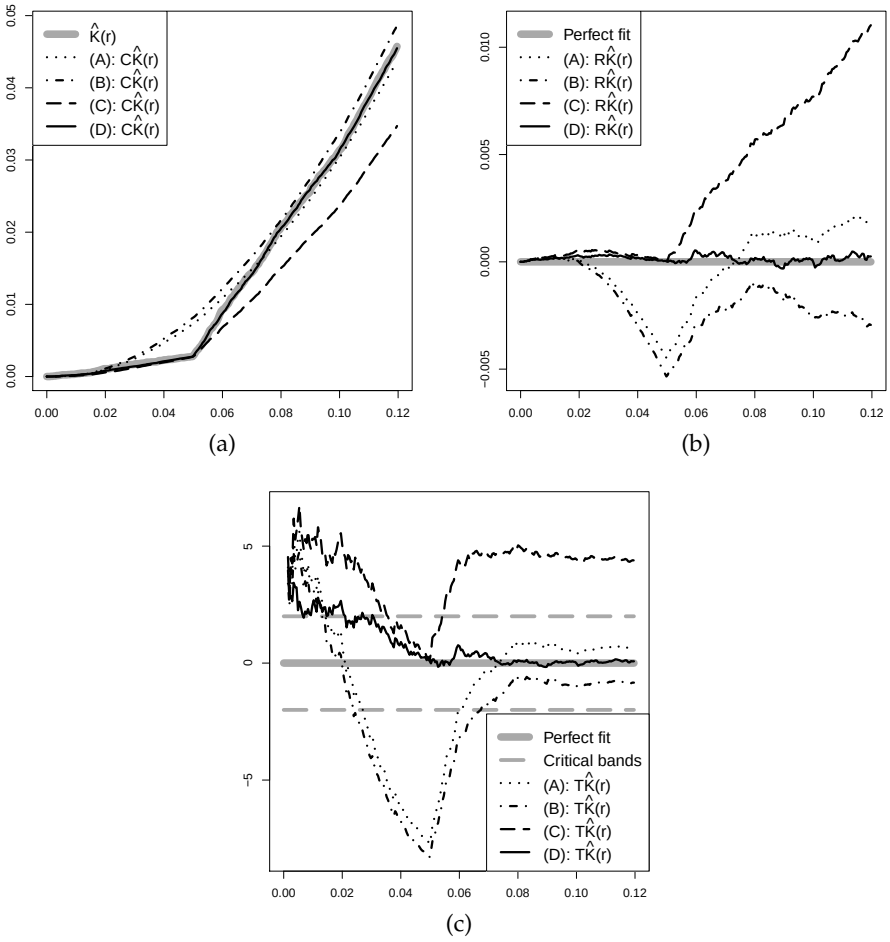
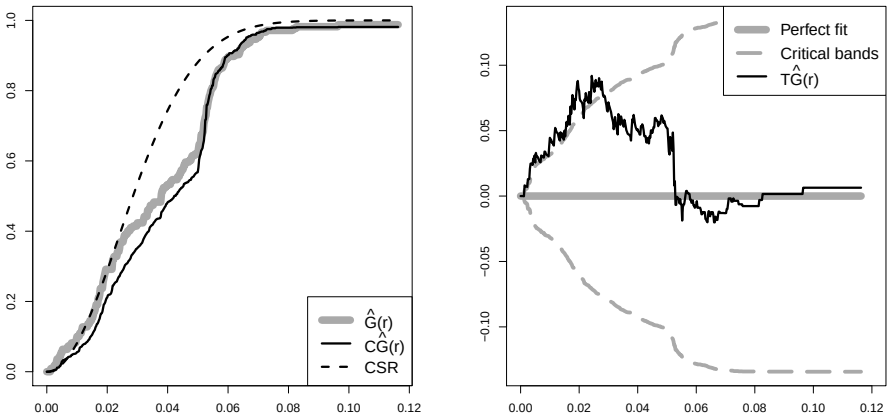


Fig 4: Goodness-of-fit diagnostics based on pairwise distances, for each of the models (A)–(D) fitted to the data in Figure 1b. (a) $\hat{K}$ and its compensator under each model. (b) residual $\hat{K}$-function (empirical minus compensator) under each model. (c) standardised residual $\hat{K}$-function under each model.

$\hat{K}$ diagnostics. This suggests that the residual of $\hat{G}$ may be particularly sensitive to lack-of-fit of interaction.



(a) (b)

Fig 5: Residual diagnostics obtained from the perturbing $\hat{G}$-model when the data pattern is a realisation of an inhomogeneous Strauss process. (a) $\hat{G}$ and its compensator under a fitted model of the correct form, and theoretical G-function for a Poisson process. (b) residual $\hat{G}$-function and two-standard-deviation limits under the fitted model of the correct form.

12.4.2 Comparison of competing models For each of the four models, Figure 6a shows $\hat{G}$ and its Papangelou compensator. This clearly shows that the Poisson models (A) and (B) fail to capture interpoint inhibition in the data. The Strauss models (C) and (D) appear virtually equivalent in Figure 6a.

Figure 6b shows the standardised residual of $\hat{G}$, and Figure 6c the pseudo-residual of V_G (i.e. the pseudo-residual based on the pertubing Geyer model), with spline smoothing applied to both plots. The Strauss models (C) and (D) appear virtually equivalent in Figure 6c. The standardised residual plot Figure 6b correctly suggests a slight lack of fit for model (C) while model (D) is judged to be a reasonable fit.

12.5 Application of $\hat{F}$ diagnostics

Figure 7 shows the pseudo-residual diagnostics based on empty space distances. Both diagnostics clearly show models (A)–(B) are poor fits to data. However, in Figure 7a it is hard to decide which of the models (C)–(D) provide a better fit. Despite the close connection between the area-interaction process and the $\hat{F}$-model, the diagnostic in Figure 7b based on the $\hat{F}$-model performs better in this particular example and correctly shows (D) is the best fit to data. In both cases it is noticed that the pseudo-sum has a much higher peak than the pseudo-compensators for the Poisson models (A)–(B), correctly suggesting that these models do not capture the strength of inhibition present in the data.

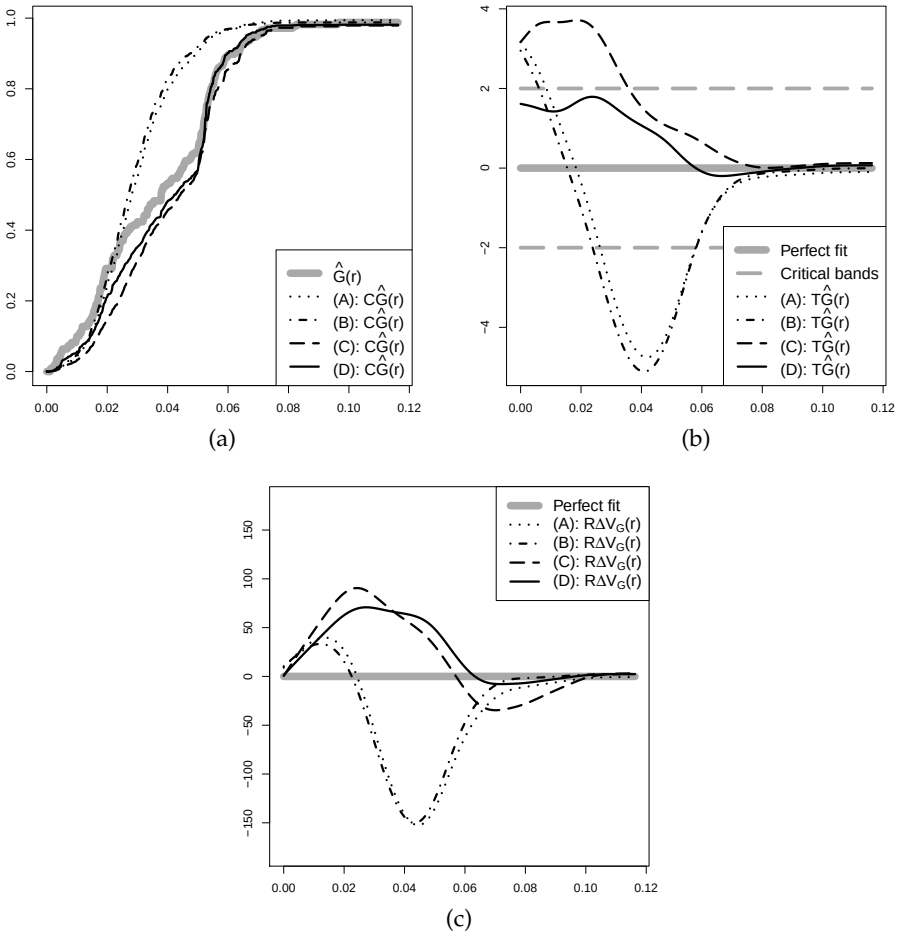


Fig 6: Diagnostics based on nearest neighbour distances, for the models (A)–(D) fitted to the data in Figure 1b. (a) compensator for $\hat{G}$. (b) smoothed standardised residual of $\hat{G}$. (c) smoothed pseudo-residual derived from a perturbing Geyer model.

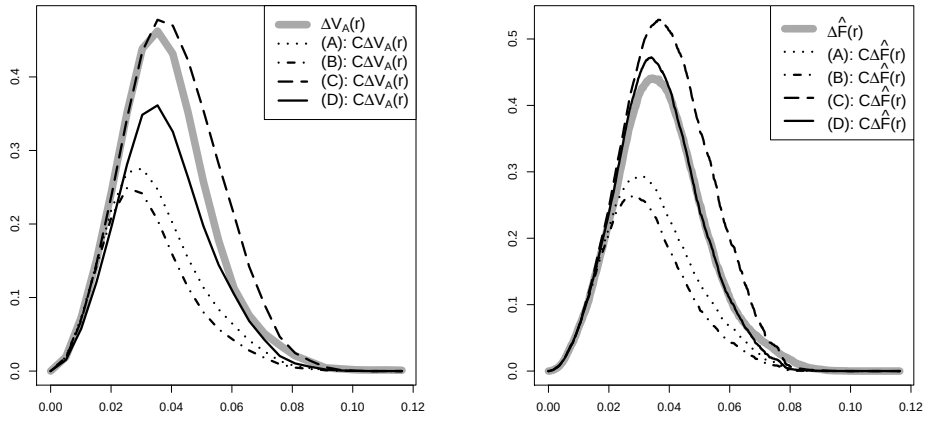


Fig 7: Pseudo-sum and pseudo-compensators for the models (A)–(D) fitted to the data in Figure 1b when the perturbing model is (a) the area-interaction process (null fitted on a fine grid) and (b) the $\hat{F}$-model (null fitted on a coarse grid).

13. TEST CASE: CLUSTERING WITHOUT TREND

13.1 Data and models

Figure 1c is a realisation of a homogeneous Geyer saturation process [30] on the unit square, with first order term $\lambda = \exp(4)$, saturation threshold $s = 4.5$ and interaction parameters $r = 0.05$ and $\gamma = \exp(0.4) \approx 1.5$, i.e. the density is

$$f(\boldsymbol{x}) \propto \exp(n(\boldsymbol{x}) \log \lambda + V_{G,s}(\boldsymbol{x}, r) \log \gamma) \tag{55}$$

where

$$V_{G,s}(\boldsymbol{x}, r) = \sum_i \min \left\{ s, \sum_{j: j \neq i} \mathbb{I}\{\|x_i - x_j\| \leq r\} \right\}.$$

This is an example of moderately strong clustering (with interaction range $R = 2r = 0.1$) without trend. The main challenge here is to correctly identify the range and type of interaction.

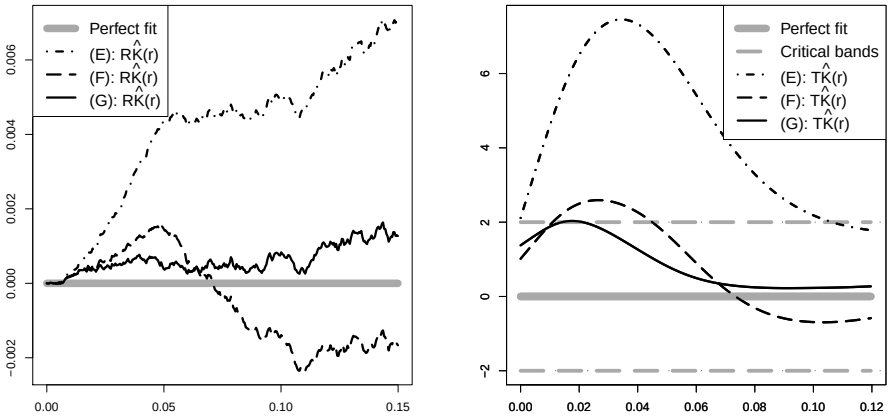
We fitted three point process models to the data: (E) a homogeneous Poisson process (CSR); (F) a homogeneous area-interaction process with disc radius $r = 0.05$; (G) a homogeneous Geyer saturation process of the correct form, with interaction parameter $r = 0.05$ and saturation threshold $s = 4.5$ while the parameters λ and γ in (55) are unknown. The parameter estimates for (G) were $\log \hat{\lambda} = 4.12$ and $\log \hat{\gamma} = 0.38$.

13.2 Application of $\hat{K}$ diagnostics

A plot (not shown) of the $\hat{K}$-function and its compensator, under each of the three models (E)–(G), demonstrates clearly that the homogeneous Poisson model (E) is a poor fit, but does not discriminate between the other models.

Figure 8 shows the residual $\hat{K}$ and the smoothed standardised residual $\hat{K}$ for the three models. These diagnostics show that the homogeneous Poisson model (E) is a poor fit, with a positive residual suggesting correctly that the data are

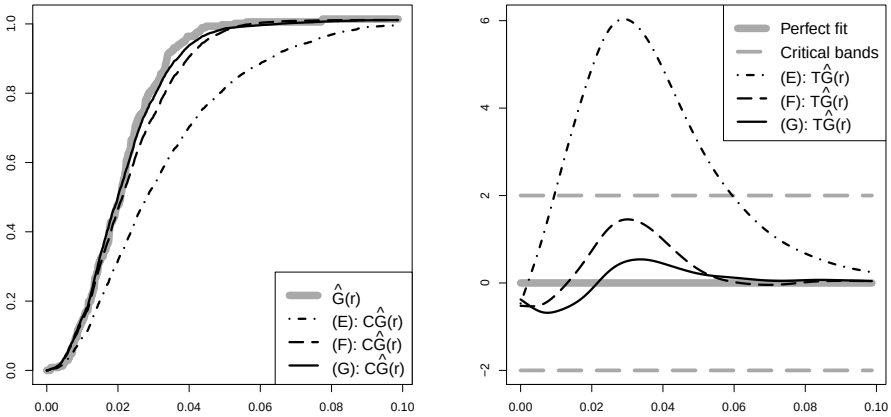
more clustered than the Poisson process. The plots suggests that both models (F) and (G) are considerably better fits to the data than a Poisson model. They show that (G) is a better fit than (F) over a range of r values, and suggest that (G) captures the correct form of the interaction.



(a) (b)

Fig 8: Goodness-of-fit diagnostics based on pairwise distances for each of the models (E)–(G) fitted to the data in Figure 1c. (a) residual $\hat{K}$; (b) smoothed standardised residual $\hat{K}$.

13.3 Application of $\hat{G}$ diagnostics



(a) (b)

Fig 9: Goodness-of-fit diagnostics based on nearest neighbour distances for each of the models (E)–(G) fitted to the data in Figure 1c. (a) $\hat{G}$ and its compensator under each model; (b) smoothed standardised residual $\hat{G}$.

Figure 9 shows $\hat{G}$ and its compensator, and the corresponding residuals and standardised residuals, for each of the models (E)–(G) fitted to the clustered point pattern in Figure 1c. The conclusions obtained from Figure 9a are the same as those in Section 13.2 based on $\hat{K}$ and its compensator. Figure 10 shows

the smoothed pseudo-residual diagnostics based on the nearest neighbour distances. The message from these diagnostics is very similar to that from Figure 9.

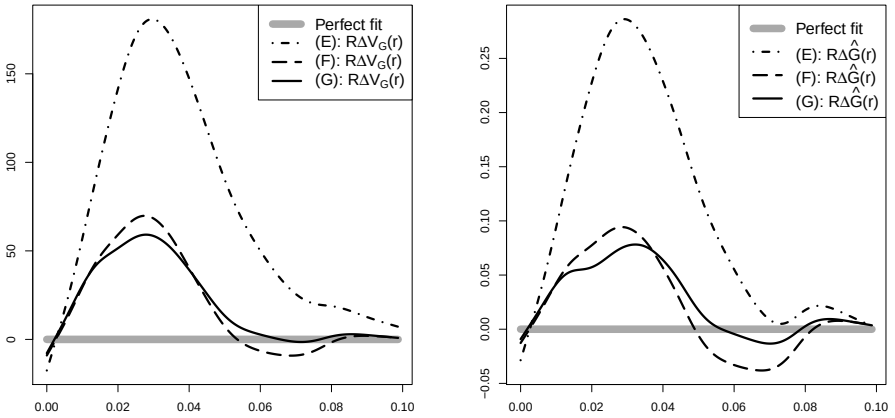


Fig 10: Smoothed pseudo-residuals for each of the models (E)–(G) fitted to the clustered point pattern in Figure 1c when the perturbing model is (a) Geyer saturation model with saturation 1, and (b) the $\hat{G}$-model.

Models (F) and (G) have the same range of interaction $R = 0.1$. Comparing Figures 8 and 9 we might conclude that the $\hat{G}$-compensator appears less sensitive to the *form* of interaction than the $\hat{K}$-compensator. Other experiments suggest that $\hat{G}$ is more sensitive than $\hat{K}$ to discrepancies in the *range* of interaction.

13.4 Application of $\hat{F}$ diagnostics

Figure 11 shows the pseudo-residual diagnostics based on the empty space distances, for the three models fitted to the clustered point pattern in Figure 1c. In this case diagnostics based on the area-interaction process and the $\hat{F}$-model are very similar, as expected due to the close connection between the two diagnostics. Here it is very noticeable that the pseudo-compensator for the Poisson model has a higher peak than the pseudo-sum, which correctly indicates that the data is more clustered than a Poisson process.

14. TEST CASE: JAPANESE PINES

14.1 Data and models

Figure 1a shows the locations of seedlings and saplings of Japanese black pine, studied by Numata [52, 53] and analysed extensively by Ogata and Tanemura [54, 55]. In their definitive analysis [55] the fitted model was an inhomogeneous 'soft core' pairwise interaction process with log-cubic first order term $\lambda_\beta(x, y) = \exp(P_\beta(x, y))$, where P_β is a cubic polynomial in x and y with coefficient vector β, and density

$$f_{(\beta,\sigma^2)}(\boldsymbol{x}) = c_{(\beta,\sigma^2)} \exp\left(\sum_i P_\beta(x_i) - \sum_{i<j} \left(\sigma^4 / \|x_i - x_j\|^4\right) \right) \tag{56}$$

where σ^2 is a positive parameter.

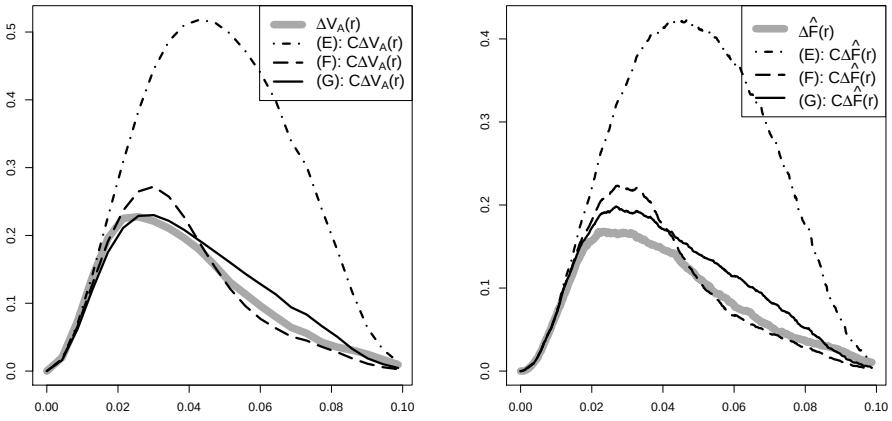


Fig 11: Pseudo-sum and pseudo-compensators for the models (E)–(G) fitted to the clustered point pattern in Figure 1c when the perturbing model is (a) area-interaction process and (b) the $\hat{F}$-model.

Here we evaluate the goodness-of-fit of three models: (H) an inhomogeneous Poisson process with log-cubic intensity; (I) a homogeneous soft core pairwise interaction process, i.e. when $P_\beta(x, y)$ in (56) is replaced by a real parameter; (J) the Ogata-Tanemura model (56). For more detail on the dataset and the fitted inhomogeneous soft core model, see [55, 6].

A complication in this case is that the soft core process (56) is not Markov, since the pair potential $c(u, v) = \exp(-\sigma^4/\|u - v\|^4)$ is always positive. Nevertheless, since this function decays rapidly, it seems reasonable to apply the residual and pseudo-residual diagnostics, using a cutoff distance R such that $|\log c(u, v)| \leq \epsilon$ when $\|u - v\| \leq R$, for a specified tolerance ϵ. The cutoff depends on the fitted parameter value σ^2. We chose $\epsilon = 0.0002$ yielding $R = 1$. Estimated interaction parameters were $\hat{\sigma}^2 = 0.11$ for model (I) and $\hat{\sigma}^2 = 0.12$ for model (J).

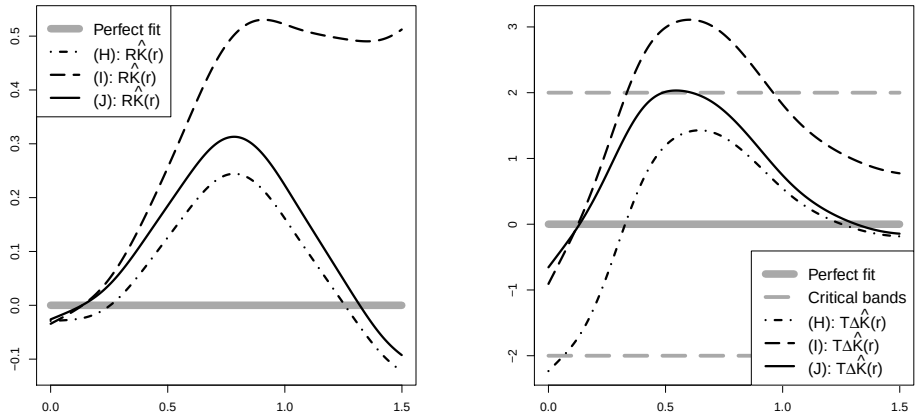
14.2 Application of $\hat{K}$ diagnostics

A plot (not shown) of $\hat{K}$ and its compensator for each of the models (H)–(J) suggests that the homogeneous soft core model (I) is inadequate, while the inhomogeneous models (H) and (J) are reasonably good fits to the data. However it does not discriminate between the models (H) and (J).

Figure 12 shows smoothed version of the residual and standardised residual of $\hat{K}$ for each model. The Ogata-Tanemura model (J) is judged to be the best fit.

14.3 Application of $\hat{G}$ diagnostics

Finally, for each of the models (H)–(J) fitted to the Japanese pines data in Figure 1a, Figure 13a shows $\hat{G}$ and its compensator. The conclusions are the same as those based on $\hat{K}$ shown in Figure 12. Figure 14 shows the pseudo-residuals when using either a perturbing Geyer model (Figure 14a) or a perturbing $\hat{G}$-model (Figure 14b). Figures 14a-14b tell almost the same story: the inhomogeneous Poisson model (H) provides the worst fit, while it is difficult to discriminate between the fit for the soft core models (I) and (J). In conclusion, consid-



(a) (b)

Fig 12: Goodness-of-fit diagnostics based on pairwise distances for each of the models (H)–(J) fitted to the Japanese pines data in Figure 1a. (a) smoothed residual $\hat{K}$; (b) smoothed standardised residual $\hat{K}$.

ering Figures 12, 13 and 14, the Ogata-Tanemura model (J) provides the best fit.

14.4 Application of $\hat{F}$ diagnostics

Finally, the empty space pseudo-residual diagnostics are shown in Figure 15 for the Japanese Pines data in Figure 1a. This gives a clear indication that the Ogata-Tanemura model (J) is the best fit to the data, and the data pattern appears to be too regular compared to the Poisson model (H) and not regular enough for the homogeneous softcore model (I).

15. SUMMARY OF TEST CASES

In this section we discuss which of the diagnostics we prefer to use based on their behaviour for the three test cases in Sections 12-14.

Typically the various diagnostics supplement each other well, and it is recommended to use more than one diagnostic when judging goodness-of-fit. Compensator and pseudo-compensator plots are informative for gaining an overall picture of goodness-of-fit, and tend to make it easy to recognize a poor fit when comparing competing models. To compare models which fit closely, it may be more informative to use (standardised) residuals or pseudo-residuals. We prefer to use the standardised residuals, but it is important not to over-interpret the significance of departure from zero.

Based on the test cases here, it is not clear whether diagnostics based on pairwise distances, nearest neighbour distances, or empty space distances are preferable. However, for each of these we prefer to work with compensators and residuals rather than pseudo-compensators and pseudo-residuals when possible (i.e. it is only necessary to use pseudo-versions for diagnostics based on empty space distances). For instance, for the first test case (Section 12) the best compensator plot is that in Figure 4a based on pairwise distances ($\hat{K}$ and C $\hat{K}$) which makes it easy to identify the correct model. On the other hand in this test case the best

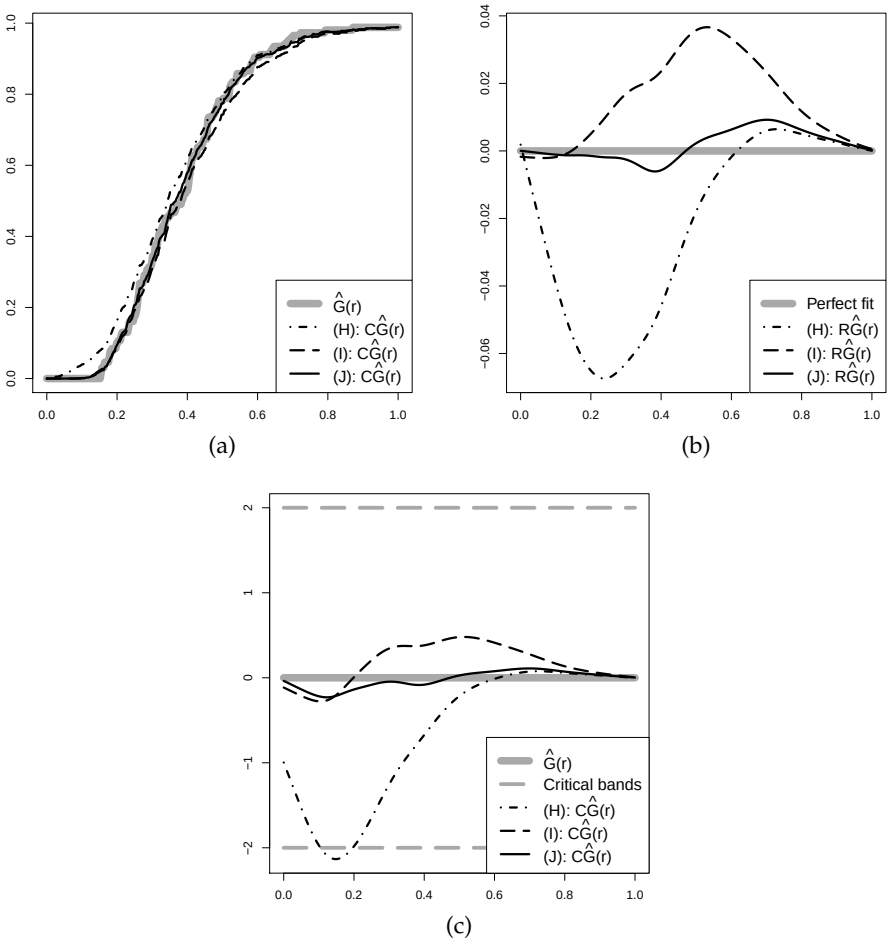
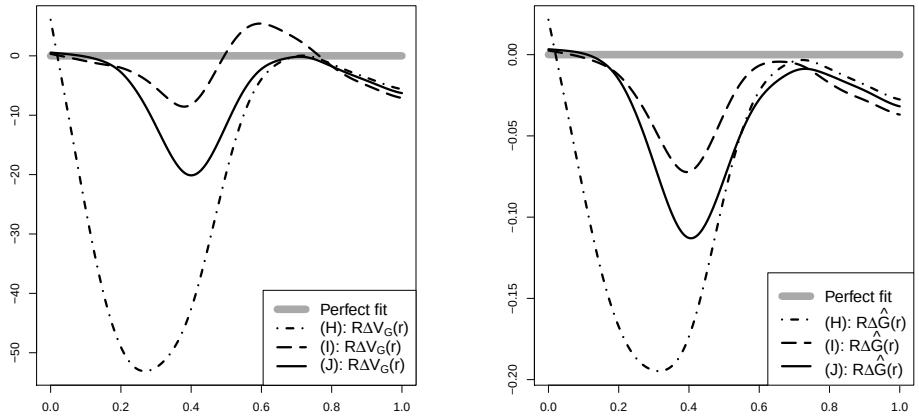
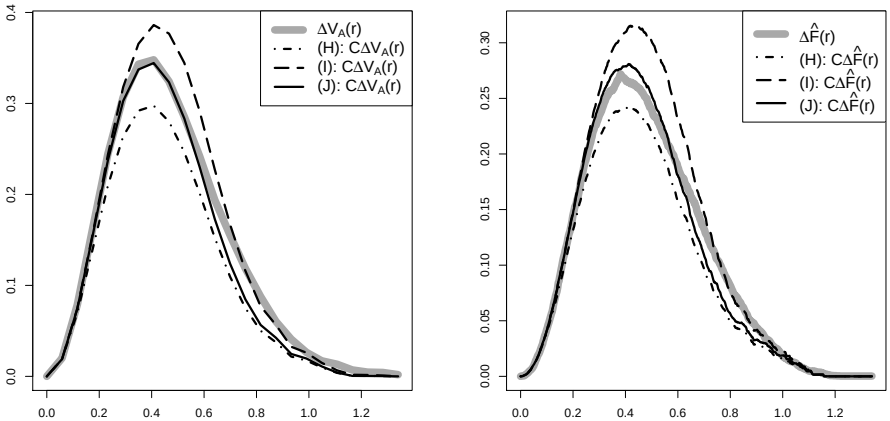


Fig 13: Goodness-of-fit diagnostics based on nearest neighbour distances for each of the models (H)–(J) fitted to the Japanese pines data in Figure 1a. (a) $\hat{G}$ and its compensator; (b) smoothed residual $\hat{G}$; (c) smoothed standardised residual $\hat{G}$.



(a) (b)

Fig 14: Smoothed pseudo-residuals for each of the models (H)–(J) fitted to the Japanese pines data in Figure 1a when the perturbing model is (a) Geyer saturation model with saturation 1 (null fitted on a fine grid) and (b) the $\hat{G}$-model.



(a) (b)

Fig 15: Pseudo-sum and pseudo-compensators for the models (H)–(J) fitted to the real data pattern in Figure 1a when the perturbing model is (a) area-interaction process and (b) the $\hat{F}$-model.

residual type plot is that in Figure 6b based on nearest neighbour distances ($\mathrm{T}\,\hat{G}$) where the correct model is the only one within the critical bands. However, in the third test case (Section 14) the best compensator plot is one of the plots in Figure 15 with pseudo-compensators based on empty space distances ($\Sigma\Delta\, V_A$ and $\mathrm{C}\Delta\, V_A$ respectively $\Sigma\Delta\,\hat{F}$ and $\mathrm{C}\Delta\,\hat{F}$) which clearly indicates which model is correct.

In the first and third test cases (Sections 12 and 14), which both involve inhomogeneous models, it is clear that $\hat{K}$ and its compensator are more sensitive to lack of fit in the first order term than $\hat{G}$ and its compensator (compare e.g. the results for the homogeneous model (C) in Figures 4a and 6a). It is our general experience that diagnostics based on $\hat{K}$ are particularly well suited to assess the presence of interaction and to identify the general form of interaction. Diagnostics based on $\hat{K}$ and in particular on $\hat{G}$ seem to be good for assessing the range of interaction.

Finally, it is worth mentioning the computational difference between the various diagnostics (timed on a 2.5 GHz laptop). The calculations for $\hat{K}$ and $\mathrm{C}\,\hat{K}$ used in Figure 2 are carried out in approximately five seconds whereas the corresponding calculations for $\hat{G}$ and $\mathrm{C}\,\hat{G}$ only take a fraction of a second. For e.g. $\Sigma\Delta\,\hat{F}$ and $\mathrm{C}\Delta\,\hat{F}$ the calculations take about 45 seconds.

ACKNOWLEDGMENTS

This paper has benefited from very fruitful discussions with Professor Rasmus P. Waagepetersen. The research was supported by The University of Western Australia, the Danish Natural Science Research Council (grants `272-06-0442` and `09-072331`, *Point process modelling and statistical inference*), the Danish Agency for Science, Technology and Innovation (grant `645-06-0528`, *International Ph.D. student*) and by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation.

REFERENCES

[1] ALM, S. E. (1988). Approximation and Simulation of the Distributions of Scan Statistics for Poisson Processes in Higher Dimensions. *Extremes* **1** 111–126.

[2] ATKINSON, A. C. (1982). Regression diagnostics, transformations and constructed variables (with discussion). *J. R. Statist. Soc. Ser. B Stat. Methodol.* **44** 1–36.

[3] BADDELEY, A., MØLLER, J. and PAKES, A. G. (2008). Properties of residuals for spatial point processes. *Ann. Inst. Statist. Math.* **60** 627–649.

[4] BADDELEY, A. and TURNER, R. (2000). Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Aust. N. Z. J. Stat.* **42** 283–322.

[5] BADDELEY, A. and TURNER, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *J. Statist. Software* **12** 1–42. URL: `www.jstatsoft.org`, ISSN: 1548-7660.

[6] BADDELEY, A., TURNER, R., MØLLER, J. and HAZELTON, M. (2005). Residual analysis for spatial point processes (with discussion). *J. R. Statist. Soc. Ser. B Stat. Methodol.* **67** 617–666.

[7] BADDELEY, A. J. (1980). A limit theorem for statistics of spatial data. *Adv. in Appl. Probab.* **12** 447–461.

[8] BADDELEY, A. J. (1993). Stereology and Survey Sampling Theory. *Bull. Int. Statist. Inst.* **50, book 2** 435–449.

[9] BADDELEY, A. J. (1999). Spatial sampling and censoring. In *Stochastic Geometry: Likelihood and Computation* (O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M. van Lieshout, eds.) 2 37–78. Chapman and Hall, London.

[10] BADDELEY, A. J. and MØLLER, J. (1989). Nearest-neighbour Markov point processes and random sets. *Int. Stat. Rev.* **57** 89–121.

[11] BADDELEY, A. J. and VAN LIESHOUT, M. N. M. (1995). Area-interaction point processes. *Ann. Inst. Statist. Math.* **47** 601–619.

[12] BARNARD, G. (1963). Contribution to discussion of "The spectral analysis of point processes" by M.S. Bartlett. *J. R. Statist. Soc. Ser. B Stat. Methodol.* **25** 294.

[13] BERMAN, M. (1986). Testing for spatial association between a point process and another stochastic process. *J. Roy. Statist. Soc. Ser. C.* **35** 54–62.

[14] BESAG, J. (1978). Some methods of statistical analysis for spatial data. *Bull. Int. Statist. Inst.* **44** 77–92.

[15] CHEN, C. (1983). Score tests for regression models. *J. Amer. Statist. Assoc.* **78** 158–161.

[16] CHETWYND, A. G. and DIGGLE, P. J. (1998). On estimating the reduced second moment measure of a stationary spatial point process. *Aust. N. Z. J. Stat.* **40** 11–15.

[17] CONNIFFE, D. (2001). Score tests when a nuisance parameter is unidentified under the null hypothesis. *J. Statist. Plann. Inference* **97** 67–83.

[18] COOK, R. D. and WEISBERG, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika* **70** 1–10.

[19] CORDY, C. B. (1993). An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statist. Probab. Lett.* **18** 353–362.

[20] COX, D. R. (1972). The statistical analysis of dependencies in point processes. In *Stochastic Point Processes* (P. A. W. Lewis, ed.) 55–66. Wiley, New York.

[21] COX, D. R. and HINKLEY, D. V. (1974). *Theoretical statistics*. Chapman and Hall, London.

[22] CRESSIE, N. A. C. (1991). *Statistics for Spatial Data*. John Wiley and Sons, New York.

[23] DALEY, D. J. and VERE-JONES, D. (1988). *An Introduction to the Theory of Point Processes*. Springer Verlag, New York.

[24] DAVIES, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64** 247–254.

[25] DAVIES, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** 33–43.

[26] DIGGLE, P. J. (1979). On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrika* **35** 87–101.

[27] DIGGLE, P. J. (1985). A kernel method for smoothing point process data. *J. Roy. Statist. Soc. Ser. C.* **34** 138–147.

[28] DIGGLE, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*, Second ed. Hodder Arnold, London.

[29] GEORGII, H. O. (1976). Canonical and grand canonical Gibbs states for continuum systems. *Comm. Math. Phys.* **48** 31–51.

[30] GEYER, C. J. (1999). Likelihood Inference for Spatial Point Processes. In *Stochastic Geometry: Likelihood and Computation*, (O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M. van Lieshout, eds.). *Monographs on Statistics and Applied Probability* 80 3 79–140. Chapman and Hall / CRC, Boca Raton, Florida.

[31] HANISCH, K. (1984). Some remarks on estimators of the distribution function of nearest neighbour distance in stationary spatial point patterns. *Math. Operationsforsch. Statist. Ser. Statist.* **15** 409–412.

[32] HANSEN, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **64** 413–430.

[33] HEINRICH, L. (1988). Asymptotic behaviour of an empirical nearest-neighbour distance function for stationary Poisson cluster processes. *Math. Nachr.* **136** 131–148.

[34] HEINRICH, L. (1988). Asymptotic Gaussianity of some estimators for reduced factorial moment measures and product densities of stationary Poisson cluster processes. *Statistics* **19** 87–106.

[35] HOPE, A. C. A. (1968). A simplified Monte Carlo significance test procedure. *J. R. Statist. Soc. Ser. B Stat. Methodol.* **30** 582–598.

[36] HUANG, F. and OGATA, Y. (1999). Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *J. Comput. Graph. Statist.* **8** 510–530.

[37] ILLIAN, J., PENTTINEN, A., STOYAN, H. and STOYAN, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley and Sons, Chichester.

[38] JENSEN, J. L. and MØLLER, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.* **1** 445–461.

[39] JOLIVET, E. (1980). Central limit theorem and convergence of empirical processes for stationary point processes. In *Point processes and queueing problems* (P. Bastfai and J. Tomko, eds.) 117–161. North-Holland, Amsterdam.

[40] KALLENBERG, O. (1978). On conditional intensities of point processes. *Z. Wahrscheinlichkeit.* **41** 205–220.

[41] KALLENBERG, O. (1984). An informal guide to the theory of conditioning in point processes. *Int. Stat. Rev.* **52** 151–164.

[42] KELLY, F. P. and RIPLEY, B. D. (1976). A note on Strauss's model for clustering. *Biometrika* **63** 357–360.

[43] KULLDORFF, M. (1999). Spatial scan statistics: models, calculations, and applications. In *Recent Advances on Scan Statistics* (J. Glaz and N. Balakrishnan, eds.) 303–322. Birkhauser, Boston.

[44] KUTOYANTS, Y. A. (1998). *Statistical Inference for Spatial Poisson Processes*. *Lecture Notes in Statistics* 134. Springer, New York.

[45] LAST, G. and PENROSE, R. (2010). Poisson process Fock space representation, chaos expansion and covariance inequalities. Manuscript in preparation.

[46] LAWSON, A. B. (1993). On the analysis of mortality events around a prespecified fixed point. *J. Roy. Statist. Soc. Ser. A* **156** 363-377.

[47] LOTWICK, H. W. and SILVERMAN, B. W. (1982). Methods for analysing spatial processes of several types of points. *J. R. Statist. Soc. Ser. B Stat. Methodol.* **44** 406–413.

[48] MØLLER, J., SYVERSVEEN, A. R. and WAAGEPETERSEN, R. (1998). Log Gaussian Cox processes. *Scand. J. Statist.* **25** 451-482.

[49] MØLLER, J. and WAAGEPETERSEN, R. P. (2007). Modern spatial point process modelling and inference (with discussion). *Scand. J. Statist.* **34** 643–711.

[50] MØLLER, J. and WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton.

[51] NGUYEN, X. X. and ZESSIN, H. (1979). Integral and differential characterizations of Gibbs processes. *Math. Nachr.* **88** 105–115.

[52] NUMATA, M. (1961). Forest vegetation in the vicinity of Choshi — coastal flora and vegetation at Choshi, Chiba prefecture, IV (in Japanese). *Bull. Choshi Mar. Lab.* 3 28–48. Chiba University.

[53] NUMATA, M. (1964). Forest vegetation, particularly pine stands in the vicinity of Choshi — flora and vegetation in Choshi, Chiba prefecture, VI (in Japanese). *Bull. Choshi Mar. Lab.* 6 27–37. Chiba University.

[54] OGATA, Y. and TANEMURA, M. (1981). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Ann. Inst. Statist. Math.* **B 33** 315–338.

[55] OGATA, Y. and TANEMURA, M. (1986). Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In *Pacific Statistical Congress* (I. S. FRANCIS, B. J. F. MANLY and F. C. LAM, eds.) 150–154. Elsevier, Amsterdam.

[56] OHSER, J. (1983). On estimators for the reduced second moment measure of point processes. *Math. Operationsforsch. Statist. Ser. Statist.* **14** 63–71.

[57] OHSER, J. and STOYAN, D. (1981). On the second-order and orientation analysis of planar stationary point processes. *Biom. J.* **23** 523-533.

[58] PAPANGELOU, F. (1974). The conditional intensity of general point processes and an application to line processes. *Z. Wahrscheinlichkeit.* **28** 207–226.

[59] PREGIBON, D. (1982). Score tests in GLIM with applications. In *GLIM 82: Proceedings of the International Conference on Generalized Linear Models*. *Lecture Notes in Statistics* 14. Springer, New York.

[60] RAO, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Math. Proc. Cambridge Philos. Soc.* **44** 50–57.

[61] RATHBUN, S. L. and CRESSIE, N. (1994). Asymptotic properties of estimators of the parameters of spatial inhomogeneous Poisson point processes. *Adv. in Appl. Probab.* **26** 122–154.

[62] RIPLEY, B. D. (1976). The second-order analysis of stationary point processes. *J. Appl. Probab.* **13** 255–266.

[63] RIPLEY, B. D. (1977). Modelling spatial patterns (with discussion). *J. R. Statist. Soc. Ser. B Stat. Methodol.* **39** 172–212.

[64] RIPLEY, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press.

[65] RIPLEY, B. D. and KELLY, F. P. (1977). Markov point processes. *J. Lond. Math. Soc.* **15** 188-192.

[66] SCHLADITZ, K. and BADDELEY, A. J. (2000). A third order point process characteristic. *Scand. J. Statist.* **27** 657–671.

[67] SILVAPULLE, M. (1996). A test in the presence of nuisance parameters. *J. Amer. Statist. Assoc.* **91** 1690–1693.

[68] STILLINGER, F. H., STILLINGER, D. K., TORQUATO, S., TRUSKETT, T. M. and DEBENEDETTI, P. G. (2000). Triangle distribution and equation of state for classical rigid disks. *J. Stat. Phys.* **100** 49–72.

[69] STOYAN, D., KENDALL, W. S. and MECKE, J. (1987). *Stochastic Geometry and its Applications*. John Wiley and Sons, Chichester.

[70] STOYAN, D. and STOYAN, H. (1995). *Fractals, Random Shapes and Point Fields*. John Wiley and Sons, Chichester.

[71] STRAUSS, D. J. (1975). A model for clustering. *Biometrika* **63** 467–475.

[72] VAN LIESHOUT, M. N. M. (2000). *Markov Point Processes and their Applications*. Imperial College Press, London.

[73] WALD, A. (1941). Some examples of asymptotically most powerful tests. *Ann. Math. Statist.* **12** 396–408.

[74] WALLER, L., TURNBULL, B., CLARK, L. C. and NASCA, P. (1992). Chronic Disease Surveillance and testing of clustering of disease and exposure: Application to leukaemia incidence and TCE-contaminated dumpsites in upstate New York. *Environmetrics* **3** 281–300.

[75] WANG, P. C. (1985). Adding a variable in generalized linear models. *Technometrics* **27** 273–276.

[76] WIDOM, B. and ROWLINSON, J. S. (1970). New model for the study of liquid-vapor phase transitions. *J. Chem. Phys.* **52** 1670–1684.

[77] WU, L. (2000). A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. *Probab. Theory Related Fields* **118** 427–438.

APPENDIX A: FURTHER DIAGNOSTICS

In this appendix we present other diagnostics which we have not implemented in software. The examples are therefore not accompanied by experimental results.

A.1 Third and higher order functional summary statistics

While the intensity and K-function are frequently-used summaries for the first and second order moment properties of a spatial point process, third and higher order summaries have been less used, though various such summaries have been suggested in e.g. [66, 48, 68, 70].

Statistic of order k For a functional summary statistic of k-th order, say

$$S(\boldsymbol{x}, r) = \sum_{\{x_{i_1},\ldots,x_{i_k}\}\subseteq \boldsymbol{x}} q(\{x_{i_1},\ldots,x_{i_k}\}, r) \tag{57}$$

we obtain

$$\Sigma\Delta\, S(\boldsymbol{x}, r) = k! S(\boldsymbol{x}, r) \; = k! \sum_{\{x_{i_1},\ldots,x_{i_k}\}\subseteq \boldsymbol{x}} q(\{x_{i_1},\ldots,x_{i_k}\}, r) \tag{58}$$

(59)

$$\mathsf{C}\Delta\, S(\boldsymbol{x}, r) = k!\,\mathsf{C}\, S(\boldsymbol{x}, r) = (k-1)! \int_W \sum_{\{x_{i_1},\ldots,x_{i_{k-1}}\}\subseteq \boldsymbol{x}} q(\{x_{i_1},\ldots,x_{i_{k-1}}, u\}, r)\lambda_{\hat{\theta}}(u, \boldsymbol{x})\,\mathrm{d}u$$

$$\mathsf{PU}(\hat{\theta}, r) = k!\,\mathrm{R}\, S(\boldsymbol{x}, r) = k! S(\boldsymbol{x}, r) - k!\,\mathsf{C}\, S(\boldsymbol{x}, r) \tag{60}$$

where $i_1, i_2, \ldots$ are pairwise distinct in the sums in (58)-(59). So in this case again, pseudo-residual diagnostics are equivalent to those based on residuals.

Third order example For a stationary and isotropic point process (i.e., when the distribution of $\boldsymbol{X}$ is invariant under translations and rotations), the intensity and K-function of the process completely determine its first and second order moment properties. However, even in this case, the simplest description of third order moments depends on a three-dimensional vector specified from triplets (x_i, x_j, x_k) of points from $\boldsymbol{X}$ such as the lengths and angle between the vectors $x_i - x_j$ and $x_j - x_k$. This is often considered too complex, and instead one considers a certain one-dimensional property of the triangle $T(x_i, x_j, x_k)$ as exemplified below, where $L(x_i, x_j, x_k)$ denotes the largest side in $T(x_i, x_j, x_k)$.

Let the canonical sufficient statistic of the perturbing density (27) be

$$S(\boldsymbol{x}, r) = V_T(\boldsymbol{x}, r) = \sum_{i<j<k} \mathbb{I}(L(x_i, x_j, x_k) \le r). \tag{61}$$

The perturbing model is a special case of the *triplet interaction point process* studied in [30]. It is also a special case of (57) with

$$q(\{x_i, x_j, x_k\}, r) = \mathbb{I}(L(x_i, x_j, x_k) \le r)$$

and so residual and pseudo-residual diagnostics are equivalent and given by (58)-(60).

A.2 Tessellation functional summary statistics

Some authors have suggested the use of tessellation methods for characterizing spatial point processes; see [37] and the references therein. A planar tessellation is a subdivision of planar region such as W or the entire plane $\mathbb{R}^2$.

For example, consider the Dirichlet tessellation of W generated by $\boldsymbol{x}$, that is, the tessellation with cells

$$C(x_i|\boldsymbol{x}) = \{u \in W | \, \|u - x_i\| \leq \|u - x_j\| \text{ for all } x_j \text{ in } \boldsymbol{x}\}, \quad i = 1, \ldots, n.$$

Suppose the canonical sufficient statistic of the perturbing density (27) is

$$S(\boldsymbol{x}, r) = V_O(\boldsymbol{x}, r) = \sum_i \mathbb{I}(|C(x_i|\boldsymbol{x})| \leq r). \tag{62}$$

This is a sum of local contributions as in (33), although not of local statistics in the sense mentioned in Section 6.3, since $\mathbb{I}(|C(x_i|\boldsymbol{x})| \leq r)$ depends on those points in $\boldsymbol{x}_{-i}$ which are Dirichlet neighbours to x_i and such points may of course not be r-close to x_i (unless r is larger than the diameter of W). We call this perturbing model for a *soft Ord's process*; Ord's process as defined in [10] is the limiting case $\phi \to -\infty$ in (27), i.e. when r is the lower bound on the size of cells. Since $V_O(\boldsymbol{x}) \leq n(\boldsymbol{x})$, the perturbing model is well-defined for all $\phi \in \mathbb{R}$.

Let $\sim_{\boldsymbol{x}}$ denote the Dirichlet neighbour relation for the points in $\boldsymbol{x}$, that is, $x_i \sim_{\boldsymbol{x}} x_j$ if $C(x_i|\boldsymbol{x}) \cap C(x_j|\boldsymbol{x}) \neq \emptyset$. Note that $x_i \sim_{\boldsymbol{x}} x_i$. Now,

$$\begin{aligned} \Delta_u S(\boldsymbol{x}, r) =& \mathbb{I}(|C(u|\boldsymbol{x} \cup \{u\})| \leq r) \\ &+ \sum_{v \neq u:\, v \sim_{\boldsymbol{x} \cup \{u\}} u} [\mathbb{I}(|C(v|\boldsymbol{x} \cup \{u\})| \leq r) - \mathbb{I}(|C(v|\boldsymbol{x} \setminus \{u\})| \leq r)] \end{aligned} \tag{63}$$

depends not only on the points in $\boldsymbol{x}$ which are Dirichlet neighbours to u (with respect to $\sim_{\boldsymbol{x} \cup \{u\}}$) but also on the Dirichlet neighbours to those points (with respect to $\sim_{\boldsymbol{x} \cup \{u\}}$ or with respect to $\sim_{\boldsymbol{x} \setminus \{u\}}$). In other words, if we define the iterated Dirichlet neighbour relation by that $x_i \sim^2_{\boldsymbol{x}} x_j$ if there exists some x_k such that $x_i \sim_{\boldsymbol{x}} x_k$ and $x_j \sim_{\boldsymbol{x}} x_k$, then $t(u, \boldsymbol{x})$ depends on those points in $\boldsymbol{x}$ which are iterated Dirichlet neighbours to u with respect to $\sim_{\boldsymbol{x} \cup \{u\}}$ or with respect to $\sim_{\boldsymbol{x} \setminus \{u\}}$. The pseudo-sum associated to the soft Ord's process is

$$\Sigma\Delta\, V_O(\boldsymbol{x}, r) = V_O(\boldsymbol{x}, r) + \sum_i \sum_{j \neq i:\, x_j \sim_{\boldsymbol{x}} x_i} [\mathbb{I}(|C(x_j|\boldsymbol{x})| \leq r) - \mathbb{I}(|C(x_j|\boldsymbol{x}_{-i})| \leq r))]$$

and from (29) and (63) we obtain the pseudo-compensator. From (36) and (62), we obtain the Papangelou compensator

$$\mathrm{C}\, V_O(\boldsymbol{x}, r) = \int_W \mathbb{I}(|C(u|\boldsymbol{x} \cup \{u\})| \leq r) \lambda_{\hat{\theta}}(u, \boldsymbol{x}) \, \mathrm{d}u.$$

Many other examples of tessellation characteristics may be of interest. For example, often the Delaunay tessellation is used instead of the Dirichlet tessellation. This is the dual tessellation to the Dirichlet tessellation, where the Delaunay cells generated by $\boldsymbol{x}$ are given by those triangles $T(x_i, x_j, x_k)$ such that the disc containing x_i, x_j, x_k in its boundary does not contain any further points from

$\boldsymbol{x}$ (strictly speaking we need to assume a regularity condition, namely that $\boldsymbol{x}$ has to be in general quadratic position; for such details, see [10]. For instance, the summary statistic $t(\boldsymbol{x}, r)$ given by the number of Delaunay cells $T(x_i, x_j, x_k)$ with $L(x_i, x_j, x_k) \leq r$ is related to (61) but concerns only the maximal cliques of Dirichlet neighbours (assuming again the general quadratic position condition). The corresponding perturbing model has to the best of our knowledge not been studied in the literature.

APPENDIX B: VARIANCE FORMULAE

This Appendix concerns the variance of diagnostic quantities of the form

$$\begin{aligned} I &= \sum_i h(x_i, \boldsymbol{X}_{-i}) - \int_W h(u, \boldsymbol{x}) \lambda_\theta(u, \boldsymbol{X}) \, \mathrm{d}u \\ R &= \sum_i h(x_i, \boldsymbol{X}_{-i}) - \int_W h(u, \boldsymbol{x}) \lambda_{\hat{\theta}}(u, \boldsymbol{X}) \, \mathrm{d}u \end{aligned}$$

where $h(\cdot)$ is a functional for which these quantities are almost surely finite, $\boldsymbol{X}$ is a point process on W with conditional intensity $\lambda_\theta(u, \boldsymbol{X})$ and $\hat{\theta}$ is an estimate of θ (e.g. the MPLE).

B.1 General identity

Exact formulae for the variance of the innovation I and residual R are given in [3]. Expressions for $\mathbb{V}\mathrm{ar}R$ are unwieldy [3, Sect. 6], but to a first approximation we may ignore the effect of estimating θ and consider the variance of I. Suppressing the dependence on θ, this is [3, Prop. 4]

$$\begin{aligned} \mathbb{V}\mathrm{ar}I &= \int_W \mathbb{E}\left[h(u, \boldsymbol{X})^2 \lambda(u, \boldsymbol{X})\right] \, \mathrm{d}u \\ &\quad + \int_{W^2} \mathbb{E}\left[A(u, v, \boldsymbol{X}) + B(u, v, \boldsymbol{X})\right] \, \mathrm{d}u \, \mathrm{d}v \end{aligned} \tag{64}$$

where

$$\begin{aligned} A(u, v, \boldsymbol{X}) &= \Delta_u h(v, \boldsymbol{X}) \, \Delta_v h(u, \boldsymbol{X}) \lambda_2(u, v, \boldsymbol{X}) \\ B(u, v, \boldsymbol{X}) &= h(u, \boldsymbol{X}) h(v, \boldsymbol{X}) \left\{\lambda_2(u, v, \boldsymbol{X}) - \lambda(u, \boldsymbol{X}) \lambda(v, \boldsymbol{X})\right\} \end{aligned}$$

where $\lambda_2(u, v, \boldsymbol{x}) = \lambda(u, \boldsymbol{x}) \lambda(v, \boldsymbol{x} \cup \{u\})$ is the second order conditional intensity. Note that for a Poisson process $B(u, v, \boldsymbol{X})$ is identically zero since $\lambda(u, \boldsymbol{X}) = \lambda(u)$.

B.2 Pseudo-score

Let $S(\boldsymbol{x}, z)$ be a functional summary statistic with function argument z. Take $h(u, \boldsymbol{X}) = \Delta_u S(\boldsymbol{x}, z)$. Then the innovation I is the pseudo-score (23), and the variance formula (64) becomes

$$\begin{aligned} \mathbb{V}\mathrm{ar}\left[\mathsf{PU}(\theta)\right] &= \int_W \mathbb{E}\left[\left(\Delta_u S(\boldsymbol{X}, z)\right)^2 \lambda(u, \boldsymbol{X})\right] \, \mathrm{d}u \\ &\quad + \int_{W^2} \mathbb{E}\left[\left(\Delta_u \Delta_v S(\boldsymbol{X}, z)\right)^2 \lambda_2(u, v, \boldsymbol{X})\right] \, \mathrm{d}u \, \mathrm{d}v \\ &\quad + \int_{W^2} \mathbb{E}\left[\Delta_u S(\boldsymbol{x}, z) \, \Delta_v S(\boldsymbol{x}, z) \left\{\lambda_2(u, v, \boldsymbol{X}) - \lambda(u, \boldsymbol{X}) \lambda(v, \boldsymbol{X})\right\}\right] \, \mathrm{d}u \, \mathrm{d}v \end{aligned} \tag{65}$$

where for $u \neq v$ and $\{u, v\} \cap \boldsymbol{x} = \emptyset$,

$$\begin{aligned}\Delta_u \Delta_v S(\boldsymbol{x}, z) &= \Delta_v \Delta_u S(\boldsymbol{x}, z) \\ &= S(\boldsymbol{x} \cup \{u, v\}, z) - S(\boldsymbol{x} \cup \{u\}, z) - S(\boldsymbol{x} \cup \{v\}, z) + S(\boldsymbol{x}, z).\end{aligned}$$

APPENDIX C: MODIFIED EDGE CORRECTIONS

Appendices C–E describe modifications to the standard edge corrected estimators of $K(r)$ and $G(r)$ that are required in the conditional case (Section 2.3) because the Papangelou conditional intensity $\lambda(u, \boldsymbol{x})$ can or should only be evaluated at locations $u \in W^\circ$ where $W^\circ \subset W$. The corresponding compensators are also given.

Assume the point process is Markov and we are in the conditional case as described in Section 5.4. Consider an empirical functional statistic of the form

$$S_W(\boldsymbol{x}, r) = \sum_{x_i \in \boldsymbol{x}} s_W(x_i, \boldsymbol{x} \setminus \{x_i\}, r) \tag{66}$$

with compensator (in the unconditional case)

$$\mathrm{C}\, S_W(\boldsymbol{x}, r) = \int_W s_W(u, \boldsymbol{x}, r) \lambda_{\hat\theta}(u, \boldsymbol{x})\, \mathrm{d}u.$$

We explore two different strategies for modifying the edge correction.

In the *restriction approach*, we replace W by W° and $\boldsymbol{x}$ by $\boldsymbol{x}^\circ = \boldsymbol{x} \cap W^\circ$ yielding

$$\begin{aligned} S_{W^\circ}(\boldsymbol{x}, r) &= \sum_{x_i \in \boldsymbol{x}^\circ} s_{W^\circ}(x_i, \boldsymbol{x}^\circ \setminus \{x_i\}, r) \\ \mathrm{C}\, S_{W^\circ}(\boldsymbol{x}, r) &= \int_{W^\circ} s_{W^\circ}(u, \boldsymbol{x}^\circ, r) \lambda_{\hat\theta}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\, \mathrm{d}u. \end{aligned} \tag{67}$$

In this approach, data points in the boundary region W^+ are ignored in the calculation of the empirical statistic S. The boundary configuration $\boldsymbol{x}^+ = \boldsymbol{x} \cap W^+$ contributes only to the estimation of θ and the calculation of the conditional intensity $\lambda_{\hat\theta}(\cdot, \cdot \mid \boldsymbol{x}^+)$. This has the advantage that the modified empirical statistic (67) is identical to the standard statistic S computed on the subdomain W°; it can be computed using existing software, and requires no new theoretical justification.

The disadvantage of the restriction approach is that we lose information by discarding some of the data. In the *reweighting approach* we retain the boundary points and compute

$$\begin{aligned} S_{W^\circ, W}(\boldsymbol{x}, r) &= \sum_{x_i \in \boldsymbol{x}^\circ} s_{W^\circ, W}(x_i, \boldsymbol{x} \setminus \{x_i\}, r) \\ \mathrm{C}\, S_{W^\circ, W}(\boldsymbol{x}, r) &= \int_{W^\circ} s_{W^\circ, W}(u, \boldsymbol{x}, r) \lambda_{\hat\theta}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\, \mathrm{d}u \end{aligned}$$

where $s_{W^\circ, W}(\cdot)$ is a modified version of $s_W(\cdot)$. Thus, boundary points contribute to the computation of the modified summary statistic $S_{W^\circ, W}$ and its compensator. The modification is designed so that $S_{W^\circ, W}$ has properties analogous to S_W.

The K-function and G-function of a point process $\boldsymbol{Y}$ in $\mathbb{R}^2$ is defined [62, 63] under the assumption that $\boldsymbol{Y}$ is second-order stationary. The standard estimators $\hat{K}_W(r)$ respectively $\hat{G}_{\boldsymbol{x}}(r)$ of the K-function and G-function are designed to be approximately pointwise unbiased estimators of $K(r)$ respectively $G(r)$ when applied to $\boldsymbol{X} = \boldsymbol{Y} \cap W$.

We do not necessarily assume stationarity, but when constructing modified summary statistics $\hat{K}_{W^\circ,W}(r)$ and $\hat{G}_{W^\circ,W}(r)$, we shall require that they are also approximately pointwise unbiased estimators of $K(r)$ respectively $G(r)$ when $\boldsymbol{Y}$ is stationary. This greatly simplifies the interpretation of plots of $\hat{K}_{W^\circ,W}(r)$ and $\hat{G}_{W^\circ,W}(r)$ and their compensators.

APPENDIX D: MODIFIED EDGE CORRECTIONS FOR THE K-FUNCTION

D.1 Horvitz-Thompson estimators

The most common nonparametric estimators of the K-function [62, 56, 9] are continuous Horvitz-Thompson type estimators [8, 19] of the form

$$\hat{K}(r) = \hat{K}_W(r) = \frac{1}{\widehat{\rho^2}(\boldsymbol{x})|W|} \sum_{i \neq j} e_W(x_i, x_j) \mathbb{I}\{\|x_i - x_j\| \leq r\}. \tag{68}$$

Here $\widehat{\rho^2} = \widehat{\rho^2}(\boldsymbol{x})$ should be an approximately unbiased estimator of the squared intensity ρ^2 for stationary processes. Usually $\widehat{\rho^2}(\boldsymbol{x}) = n(n-1)/|W|^2$ where $n = n(\boldsymbol{x})$.

The term $e_W(u, v)$ is an edge correction weight, depending on the geometry of W, designed so that the double sum in (68), say $\hat{Y}(r) = \widehat{\rho^2}(\boldsymbol{x})|W|\hat{K}(r)$, is an unbiased estimator of $Y(r) = \rho^2|W|K(r)$. Popular examples are the Ohser-Stoyan translation edge correction with

$$e_W(u, v) = e_W^{\text{trans}}(u, v) = \frac{|W|}{|W \cap (W + (u - v))|} \tag{69}$$

and Ripley's isotropic correction with

$$e_W(u, v) = e_W^{\text{iso}}(u, v) = \frac{2\pi||u - v||}{\mathsf{length}(\partial B(u, ||u - v||) \cap W)}. \tag{70}$$

Estimators of the form (68) satisfy the local decomposition (66) where

$$s_W(u, \boldsymbol{x}, r) = \frac{1}{\widehat{\rho^2}(\boldsymbol{x} \cup \{u\})|W|} \sum_j e_W(u, x_j) \mathbb{I}\{||u - x_j|| \leq r\}, \quad u \notin \boldsymbol{x}.$$

Now we wish to modify (68) so that the outer summation is restricted to data points x_i in $W^\circ \subset W$, while retaining the property of unbiasedness for stationary and isotropic point processes.

The *restriction estimator* is

$$\hat{K}_{W^\circ}(r) = \frac{1}{\widehat{\rho^2}(\boldsymbol{x}^\circ)|W^\circ|} \sum_{x_i \in \boldsymbol{x}^\circ} \sum_{x_j \in \boldsymbol{x}^\circ_{-i}} e_{W^\circ}(x_i, x_j) \mathbb{I}\{\|x_i - x_j\| \leq r\} \tag{71}$$

where the edge correction weight is given by (69) or (70) with W replaced by W°.

A more efficient alternative is to replace (68) by the *reweighting estimator*

$$\hat{K}_{W^\circ,W}(r) = \frac{1}{\widehat{\rho^2}(\boldsymbol{x})|W^\circ|} \sum_{x_i \in \boldsymbol{x}^\circ} \sum_{x_j \in \boldsymbol{x}_{-i}} e_{W^\circ,W}(x_i, x_j) \mathbb{I}\{||x_i - x_j|| \le r\} \tag{72}$$

where $e_{W^\circ,W}(u,v)$ is a modified version of $e_W(\cdot)$ constructed so that the double sum in (72) is unbiased for $Y(r)$. Compared to the restriction estimator (71), the reweighting estimator (72) contains additional contributions from point pairs (x_i, x_j) where $x_i \in \boldsymbol{x}^\circ$ and $x_j \in \boldsymbol{x}^+$.

The modified edge correction factor $e_{W^\circ,W}(\cdot)$ for (72) is the Horvitz-Thompson weight [9] in an appropriate sampling context. Ripley's [62, 63] isotropic correction (70) is derived assuming isotropy, by Palm conditioning on the location of the first point x_i, and determining the probability that x_j would be observed inside W after a random rotation about x_i. Since the constraint on x_j is unchanged, no modification of the edge correction weight is required, and we take $e_{W^\circ,W}(\cdot) = e_W(\cdot)$ as in (70). Note however that the denominator in (72) is changed from $|W|$ to $|W^\circ|$.

The Ohser-Stoyan [57] translation correction (69) is derived by considering two-point sets (x_i, x_j) sampled under the constraint that both x_i and x_j are inside W. Under the modified constraint that $x_i \in W^\circ$ and $x_j \in W$, the appropriate edge correction weight is

$$e_{W^\circ,W}(u,v) = e_{W^\circ,W}(u-v) = \frac{|W \cap (W^\circ + (u-v))|}{|W^\circ|}$$

so that $1/e_{W^\circ,W}(z)$ is the fraction of locations u in W° such that $u + z \in W$.

D.2 Border correction

A slightly different creature is the border corrected estimator (using usual intensity estimator $\hat{\rho} = n(\boldsymbol{x})/|W|$)

$$\hat{K}_W(r) = \frac{|W|}{n(\boldsymbol{x}) n(\boldsymbol{x} \cap W_{\ominus r})} \sum_{x_i \in \boldsymbol{x}} \sum_{x_j \in \boldsymbol{x}_{-i}} \mathbb{I}\{x_i \in W_{\ominus r}\} \mathbb{I}\{||x_i - x_j|| \le r\}$$

with compensator (in the unconditional case)

$$\mathrm{C}\,\hat{K}_W(r) = \int_{W_{\ominus r}} \frac{|W| \sum_{x_j \in \boldsymbol{x}} \mathbb{I}\{||u - x_j|| \le r\}}{(n(\boldsymbol{x})+1)(n(\boldsymbol{x} \cap W_{\ominus r})+1)} \lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\, \mathrm{d}u.$$

The restriction estimator is

$$\hat{K}_{W^\circ}(r) = \frac{|W^\circ|}{n(\boldsymbol{x}^\circ) n(\boldsymbol{x} \cap W^\circ_{\ominus r})} \sum_{x_i \in \boldsymbol{x}^\circ} \sum_{x_j \in \boldsymbol{x}^\circ_{-i}} \mathbb{I}\{x_i \in W^\circ_{\ominus r}\} \mathbb{I}\{||x_i - x_j|| \le r\}$$

and the compensator is

$$\mathrm{C}\,\hat{K}_{W^\circ}(r) = \int_{W^\circ_{\ominus r}} \frac{|W^\circ| \sum_{x_j \in \boldsymbol{x}^\circ} \mathbb{I}\{||u - x_j|| \le r\}}{(n(\boldsymbol{x}^\circ)+1)(n(\boldsymbol{x} \cap W^\circ_{\ominus r})+1)} \lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\, \mathrm{d}u.$$

Typically, $W^\circ = W_{\ominus R}$, so $W^\circ_{\ominus r}$ is equal to $W_{\ominus(R+r)}$.

The *reweighting estimator* is

$$\hat{K}_{W^\circ,W}(r) = \frac{|W|}{n(\boldsymbol{x})n(\boldsymbol{x}^\circ \cap W_{\ominus r})} \sum_{x_i \in \boldsymbol{x}^\circ} \sum_{x_j \in \boldsymbol{x}_{-i}} \mathbb{I}\{x_i \in W_{\ominus r}\}\mathbb{I}\{||x_i - x_j|| \le r\}$$

and the compensator is

$$\mathrm{C}\,\hat{K}_{W^\circ,W}(r) = \int_{W^\circ \cap W_{\ominus r}} \frac{|W| \sum_{x_j \in \boldsymbol{x}} \mathbb{I}\{||u - x_j|| \le r\}}{(n(\boldsymbol{x})+1)(n(\boldsymbol{x}^\circ \cap W_{\ominus r})+1)} \lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\,\mathrm{d}u.$$

Usually $W^\circ = W_{\ominus R}$, so $W^\circ \cap W_{\ominus r}$ is equal to $W_{\ominus \max(R,r)}$. From this we conclude that when using border correction we should always use the reweighting estimator since the restriction estimator discards additional information and neither the implementation nor the interpretation is easier.

APPENDIX E: MODIFIED EDGE CORRECTIONS FOR NEAREST NEIGHBOUR FUNCTION G

E.1 Hanisch estimators

Hanisch [31] considered estimators for $G(r)$ of the form $\hat{G}_W(r) = \hat{D}_{\boldsymbol{x}}(r)/\hat{\rho}$, where $\hat{\rho}$ is some estimator of the intensity ρ, and

$$\hat{D}_{\boldsymbol{x}}(r) = \sum_{x_i \in \boldsymbol{x}} \frac{\mathbb{I}\{x_i \in W_{\ominus d_i}\}\mathbb{I}\{d_i \le r\}}{|W_{\ominus d_i}|} \tag{73}$$

where $d_i = d(x_i, \boldsymbol{x} \setminus \{x_i\})$ is the nearest neighbour distance for x_i. If $\hat{\rho}$ were replaced by ρ then $\hat{G}_W(r)$ would be an unbiased, Horvitz-Thompson estimator of $G(r)$. See [69, pp. 128–129], [9].

Hanisch's recommended estimator D_4 is the one in which $\hat{\rho}$ is taken to be

$$\hat{D}_{\boldsymbol{x}}(\infty) = \sum_{x_i \in \boldsymbol{x}} \frac{\mathbb{I}\{x_i \in W_{\ominus d_i}\}}{|W_{\ominus d_i}|}.$$

This is sensible because $\hat{D}_{\boldsymbol{x}}(\infty)$ is an unbiased estimator of ρ and is positively correlated with $\hat{D}_{\boldsymbol{x}}(r)$. The resulting estimator $\hat{G}_W(r)$ can be decomposed in the form (66) where

$$s_W(u, \boldsymbol{x}, r) = \frac{\mathbb{I}\{u \in W_{\ominus d(u,\boldsymbol{x})}\}\mathbb{I}\{d(u,\boldsymbol{x}) \le r\}}{\hat{D}_{\boldsymbol{x}\cup\{u\}}(\infty)|W_{\ominus d(u,\boldsymbol{x})}|}$$

for $u \notin \boldsymbol{x}$, where $d(u, \boldsymbol{x})$ is the ('empty space') distance from location u to the nearest point of $\boldsymbol{x}$. Hence the corresponding compensator is

$$\mathrm{C}\,\hat{G}_W(r) = \int_W \frac{\mathbb{I}\{u \in W_{\ominus d(u,\boldsymbol{x})}\}\mathbb{I}\{d(u,\boldsymbol{x}) \le r\}}{\hat{D}_{\boldsymbol{x}\cup\{u\}}(\infty)|W_{\ominus d(u,\boldsymbol{x})}|} \lambda_{\hat{\theta}}(u, \boldsymbol{x})\,\mathrm{d}u$$

This is difficult to evaluate, since the denominator of the integrand involves a summation over all data points: $D_{\boldsymbol{x}\cup\{u\}}(\infty)$ is not related in a simple way to $D_{\boldsymbol{x}}(\infty)$.

Instead, we choose $\hat{\rho}$ to be the conventional estimator $\hat{\rho} = n(\boldsymbol{x})/|W|$. Then

$$\hat{G}_W(r) = \frac{|W|}{n(\boldsymbol{x})} \hat{D}_{\boldsymbol{x}}(r)$$

which can be decomposed in the form (66) with

$$s_W(u, \boldsymbol{x}, r) = \frac{|W|}{n(\boldsymbol{x})+1} \frac{\mathbb{I}\{u \in W_{\ominus d(u,\boldsymbol{x})}\}\mathbb{I}\{d(u,\boldsymbol{x}) \le r\}}{|W_{\ominus d(u,\boldsymbol{x})}|}$$

for $u \notin \boldsymbol{x}$, so that the compensator is

$$\mathrm{C}\hat{G}_W(r) = \frac{|W|}{n(\boldsymbol{x})+1} \int_W \frac{\mathbb{I}\{u \in W_{\ominus d(u,\boldsymbol{x})}\}\mathbb{I}\{d(u,\boldsymbol{x}) \le r\}}{|W_{\ominus d(u,\boldsymbol{x})}|} \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u. \tag{74}$$

In the *restriction estimator* we exclude the boundary points and take $d_i^\circ = d(x_i, \boldsymbol{x}^\circ_{-i})$, effectively replacing the dataset $\boldsymbol{x}$ by its restriction $\boldsymbol{x}^\circ = \boldsymbol{x} \cap W^\circ$.

$$\hat{G}_{W^\circ}(r) = \frac{|W^\circ|}{n(\boldsymbol{x}^\circ)} \sum_{x_i \in \boldsymbol{x}^\circ} \frac{\mathbb{I}\{x_i \in W^\circ_{\ominus d_i^\circ}\}\mathbb{I}\{d_i^\circ \le r\}}{|W^\circ_{\ominus d_i^\circ}|}$$

The compensator is (74) but computed for the point pattern $\boldsymbol{x}^\circ$ in the window W°:

$$\mathrm{C}\hat{G}_{W^\circ}(r) = \frac{|W^\circ|}{n(\boldsymbol{x}^\circ)+1} \int_{W^\circ} \frac{\mathbb{I}\{u \in W^\circ_{\ominus d(u,\boldsymbol{x}^\circ)}\}\mathbb{I}\{d(u,\boldsymbol{x}^\circ) \le r\}}{|W^\circ_{\ominus d(u,\boldsymbol{x}^\circ)}|} \lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\, \mathrm{d}u.$$

In the usual case $W^\circ = W_{\ominus R}$, we have $W^\circ_{\ominus d} = W_{\ominus(R+d)}$.

In the *reweighting estimator* we take $d_i = d(x_i, \boldsymbol{x} \setminus \{x_i\})$. To retain the Horvitz-Thompson property we must replace the weights $1/|W_{\ominus d_i}|$ in (73) by $1/|W^\circ \cap W_{\ominus d_i}|$. Thus the modified statistics are

$$\hat{G}_{W^\circ,W}(r) = \frac{|W|}{n(\boldsymbol{x})} \sum_{x_i \in \boldsymbol{x}^\circ} \frac{\mathbb{I}\{x_i \in W_{\ominus d_i}\}\mathbb{I}\{d_i \le r\}}{|W^\circ \cap W_{\ominus d_i}|} \tag{75}$$

and

$$\mathrm{C}\hat{G}_{W^\circ,W}(r) = \frac{|W|}{n(\boldsymbol{x})+1} \int_{W^\circ} \frac{\mathbb{I}\{u \in W_{\ominus d(u,\boldsymbol{x})}\}\mathbb{I}\{d(u,\boldsymbol{x}) \le r\}}{|W^\circ \cap W_{\ominus d(u,\boldsymbol{x})}|} \lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\, \mathrm{d}u. \tag{76}$$

In the usual case where $W^\circ = W_{\ominus R}$ we have $W^\circ \cap W_{\ominus d_i} = W_{\ominus \max(R,d_i)}$.

Optionally we may also replace $|W|/n(\boldsymbol{x})$ in (75) by $|W^\circ|/n(\boldsymbol{x} \cap W^\circ)$, and correspondingly replace $\frac{|W|}{n(\boldsymbol{x})+1}$ in (76) by $|W^\circ|/(n(\boldsymbol{x} \cap W^\circ)+1)$.

E.2 Border correction

The classical border correction estimate of G is

$$\hat{G}_W(r) = \frac{1}{n(\boldsymbol{x} \cap W_{\ominus r})} \sum_{x_i \in \boldsymbol{x}} \mathbb{I}\{x_i \in W_{\ominus r}\}\mathbb{I}\{d(x_i, \boldsymbol{x}_{-i}) \le r\} \tag{77}$$

with compensator (in the unconditional case)

$$\mathrm{C}\hat{G}_W(r) = \frac{1}{1 + n(\boldsymbol{x} \cap W_{\ominus r})} \int_{W_{\ominus r}} \mathbb{I}\{d(u,\boldsymbol{x}) \le r\} \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u. \tag{78}$$

In the conditional case, the Papangelou conditional intensity $\lambda_{\hat{\theta}}(u, \boldsymbol{x})$ must be replaced by $\lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)$ given in (24). The *restriction estimator* is obtained by replacing W by W° and $\boldsymbol{x}$ by $\boldsymbol{x}^\circ$ in (77)–(78) yielding

$$\begin{aligned} \hat{G}_{W^\circ}(r) &= \frac{1}{n(\boldsymbol{x} \cap W^\circ_{\ominus r})} \sum_{x_i \in \boldsymbol{x}^\circ} \mathbb{I}\{x_i \in W^\circ_{\ominus r}\} \mathbb{I}\{d(x_i, \boldsymbol{x}^\circ_{-i}) \le r\} \\ \mathrm{C}\, \hat{G}_{W^\circ}(r) &= \frac{1}{1 + n(\boldsymbol{x} \cap W^\circ_{\ominus r})} \int_{W^\circ_{\ominus r}} \mathbb{I}\{d(u, \boldsymbol{x}^\circ) \le r\} \lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\, \mathrm{d}u. \end{aligned}$$

Typically $W^\circ = W_{\ominus R}$ so that $W^\circ_{\ominus r} = W_{\ominus(R+r)}$. The *reweighting estimator* is obtained by restricting x_i and u in (77)–(78) to lie in W°, yielding

$$\begin{aligned} \hat{G}_{W^\circ, W}(r) &= \frac{1}{n(\boldsymbol{x}^\circ \cap W_{\ominus r})} \sum_{x_i \in \boldsymbol{x}^\circ} \mathbb{I}\{x_i \in W_{\ominus r}\} \mathbb{I}\{d(x_i, \boldsymbol{x}_{-i}) \le r\} \\ \mathrm{C}\, \hat{G}_{W^\circ, W}(r) &= \frac{1}{1 + n(\boldsymbol{x}^\circ \cap W_{\ominus r})} \int_{W^\circ \cap W_{\ominus r}} \mathbb{I}\{d(u, \boldsymbol{x}) \le r\} \lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\, \mathrm{d}u. \end{aligned}$$

In the usual case where $W^\circ = W_{\ominus R}$ we have $W^\circ \cap W_{\ominus r} = W_{\ominus \max(R,r)}$.

In the same way as for the border corrected estimate for the K-function we always choose to use the reweighting estimator rather than the restriction estimator since there are no disadvantages connected with this.

The border corrected estimator $\hat{G}(r)$ is well known for having relatively poor performance and sample properties. It is not guaranteed to be a monotonically increasing function of r, and its bias and variance are generally greater than those of the Horvitz-Thompson style estimators. The main reason for choosing the border corrected estimator is its computational efficiency in large datasets. We may expect similar considerations to apply to its compensator.